

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM protein - protein search, using sw model

Run on: August 12, 2002, 10:12:13 ; Search time 27.45 Seconds  
(Without alignments)  
1153.300 Million cell updates/sec

Title: Perfect score: US-10-080-797-1

Sequence: 1 HSHRFQPVHLVALNSPLS.....SCHHAYIVICLIENTSFMTASK 183

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172949429 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 19,\*

1: sp\_archea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query	Length	DB ID	Description
1	840	B6..9	184	11 Q9JK63	Q9jk63 mus musculu
2	840	B6..9	1140	11 061434	061434 mus musculu
3	840	B6..9	1774	11 062001	062001 mus musculu
4	820	B4..8	226	11 Q9QZD2	09qzd2 ratetus norv
5	777	B0..4	171	11 Q9WUW5	Q9wuw5 ratetus norv
6	771	B9..7	1344	13 Q93419	Q93419 gallus gall
7	722	B4..7	160	11 Q9CRT2	Q9cr2 mus musculu
8	544	B6..3	1367	11 Q9ZB06	035206 mus musculu
9	541	B5..9	1367	11 Q9EQD9	09eqd9 mus musculu
10	530	B4..8	1388	4 Q9Y4W4	Q9y4w4 homo sapien
11	529	B4..5	102	4 Q96T70	Q96t70 homo sapien
12	430.5	B4..5	581	09V509	09v509 drosophila
13	373	B8..6	650	5 017866	017866 caenorhabdi
14	373	B8..6	778	5 09U9K6	09u9k6 caenorhabdi
15	373	B8..6	864	5 093336	093336 caenorhabdi
16	373	B8..6	1117	5 09u9K7	09u9k7 caenorhabdi

PRED. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**ALIGNMENTS**

RESULT	ID	PRELIMINARY;	PRT;	184 AA.
09JK63	09JK63	09jk63; 15, Created)	DT	01-OCT-2000 (TREMBREL 15, Last sequence update)
09JK63	09JK63	09jk63; 15, Last annotation update)	DT	01-OCT-2000 (TREMBREL 15, Last sequence update)
09JK63	09JK63	09jk63; 15, Last annotation update)	DT	01-DEC-2001 (TREMBREL 19, Last annotation update)
09JK63	09JK63	09jk63; 15, Last annotation update)	DE	ENDOSTATIN (FRAGMENT).
09JK63	09JK63	09jk63; 15, Last annotation update)	OS	MUS musculus (Mouse);
09JK63	09JK63	09jk63; 15, Last annotation update)	OS	Eukaryota; Metazoa; Chordata; Cniana; Vertebrata; Euteleostomi; Mammalia; Etherida; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
09JK63	09JK63	09jk63; 15, Last annotation update)	OC	Mammalia; Etherida; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
09JK63	09JK63	09jk63; 15, Last annotation update)	OX	NBBI_TAXID=10090;
09JK63	09JK63	09jk63; 15, Last annotation update)	RN	[1]
09JK63	09JK63	09jk63; 15, Last annotation update)	RP	SEQUENCE FROM N.A.
09JK63	09JK63	09jk63; 15, Last annotation update)	RC	RC STRAIN=CHINESE KUNMING;
09JK63	09JK63	09jk63; 15, Last annotation update)	RA	RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;
09JK63	09JK63	09jk63; 15, Last annotation update)	RT	RT "Anticancer treatment of targeted fusion protein delivery to tumor neovasculature";
09JK63	09JK63	09jk63; 15, Last annotation update)	RL	RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
09JK63	09JK63	09jk63; 15, Last annotation update)	DR	DR EMBL: AF257775; AAF69009.1; -.
09JK63	09JK63	09jk63; 15, Last annotation update)	DR	DR RSPB; P39061; 1KB.
09JK63	09JK63	09jk63; 15, Last annotation update)	FT	FT NON_TER 1 1
09JK63	09JK63	09jk63; 15, Last annotation update)	FT	FT NON_TER 184 184
09JK63	09JK63	09jk63; 15, Last annotation update)	FT	FT NON_TER 184 AA; 20376 MW; AC06F9D8D103412A CRC64;
09JK63	09JK63	09jk63; 15, Last annotation update)	SQ	SQ SEQUENCE 184 AA;

Query Match Best Local Similarity 85.6%; Pred. No. 2.4e-78; Score 840; DB 11; Length 184; Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0; Gaps 0;

Qy 1 HSHRFQPVHLVALNSPLSPLSGMGRGIRGADFOCFCQDQARAVLAGTRAFRAFLSSRLQIYI 60  
Db 1 HSHRFQPVHLVALNSPLSPLSGMGRGIRGADFOCFCQDQARAVLAGTRAFRAFLSSRLQIYI 60  
Db 1 HTHQDOPVLHVALNPLPSLGMGRGIRGADFOCFCQDQARAVLAGTRAFRAFLSSRLQIYI 60

Qy 61 VRRADRAVPTVNLKDELFLPFWEAFLFGSGEPLKPGARIFSDGKDVLRLRHTWPQKSVW 120  
Db 61 VRRADRAVPTVNLKDELFLPFWEAFLFGSGEPLKPGARIFSDGKDVLRLRHTWPQKSVW 120  
Qy 121 HGSDPGRRLTESCYTERTEAPSATQASSLIGGLIGLQASASCHAYIVICLIENTSFMT 180

Db	121	HGSDPSGRRLMESCYETWRTEGTATGQASSLLSGRLEQKAASCHNSYIVLCIENSFT	180	RC TISSUE="PRIMARY TAIL CULTURE"; RX MEDLINE=94245707; PubMed=8188673;
Oy	181 A 181			RA Rehn M., Hintikka E., Pihlajaniemi T.;
Db	181 S 181			RT "Primary structure of the alpha 1 chain of mouse type XVII collagen, partial structure of the corresponding gene, and comparison of the RT alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen chain"; J. Biol. Chem. 269:13929-13935(1994).
RESULT	2			RL J. Biol. Chem. 269:13929-13935(1994).
061434		PRELIMINARY;	PRT; 1140 AA.	RN [2]
ID	061434			RN SEQUENCE OF 1-552 FROM N.A.
AC	061434			RX MEDLINE=94240112; PubMed=8183894;
DT	01-NOV-1996 (TREMBLrel. 01, Created)			RA Rehn M., Pihlajaniemi T.;
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			RT "Alpha 1(XVII), a collagen chain with frequent interruptions in the RT collagenous sequence, a distinct tissue distribution, and homology with type XV collagen";
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			RT with type XV collagen";
DE	COLLAGEN (FRAGMENT).			RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
GN	COL15A1.			RN [3]
OS	Mus musculus (Mouse)			RN SEQUENCE OF 1-562 FROM N.A.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eureleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			RX MEDLINE=95181468; PubMed=7876242;
OX	NCBI_TaxID=10090;			RA Rehn M., Pihlajaniemi T.
RN	[1]			RT "Identification of a novel collagen chain represented by extensive RT interruptions in the triple-helical region.;"
RP	SEQUENCE FROM N.A.			RT Cell. Mol. Biol. Res. 196:576-582(1993).
RA	Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;			RL EMBL; D1746; IKA0403.1; -.
RT	"Identification of a novel collagen chain represented by extensive RT interruptions in the triple-helical region.;"			DR EMBL; U03715; AAC52903.1; JOINED.
RL	Cell. Mol. Biol. Res. 196:576-582(1993).			DR EMBL; U03718; AAC52903.1; JOINED.
DR	HSSP: P3061; IKA0403.1; -.			DR EMBL; U34607; AAC52903.1; JOINED.
DR	MGD; MGI:88449; Collsal.			DR EMBL; U34609; AAC52903.1; JOINED.
DR	Interpro; IPR000087; Collagen.			DR EMBL; U34610; AAC52903.1; JOINED.
DR	PFAM: PF01391; Collagen; 6.			DR EMBL; U34611; AAC52903.1; JOINED.
FT	NON_TER 1			DR EMBL; U34612; AAC52903.1; JOINED.
SQ	SEQUENCE 1140 AA; 115156 MW: 880C7E6862B3BDFE CRC64;			DR EMBL; U11637; AAC52179.1; -.
Query	Match	86.9%	Score 840; DB 11; Length 1140;	DR HSSP; P3061; IKA0403.1; -.
Best	Local Similarity	85.6%	Pred. No. 2, 6e-77; Mismatches 11; Indels 0; Gaps 0;	DR MGD; MGI:88451; Collsal.
Matches	155; Conservative			DR Interpro; IPR000087; Collagen.
Oy	1	HSHRDQPVLPVLVALNPLSPSGMGRGTRGADFOCFOARAVGLGTRAFATSSRQLDLYSTI 60		DR InterPro; IPR00024; Fz_domain.
Db	957	HTHQDQPVLPVLVALNPLSPSGMGRGTRGADFOCFOARAVGLGTRAFATSSRQLDLYSTI 1016		DR InterPro; IPR003129; TSPN:
Oy	61	VRRADRAAVPVNLKDELFLPSWEALFSGSGPDKPGARIFSFDSKGDKVLRHPTWQOKSWV 120		DR PFAM; PF01392; FZ; 1.
Db	1017	VRRADRGSPVTPNLKDELFLPSWDLSLFSQGQLOQGPARTSFSDGRDVLRHPTWQOKSWV 1076		DR PFAM; PF02210; TSPN; 1.
Oy	121	HGSDPGRRLTESYCETWRTEPATGQASSLLGGLGQASCHAYIVLCIENSFT 180		DR SMART; SM00063; FRI; 1.
Db	1077	HGSQPSGRRLMESYCETWRTEPATGQASSLLGGLGQASCHAYIVLCIENSFT 1136		DR SMART; SM00282; LamG; 1.
Query	181 A 181			DR PROSITE; PS50038; FZ; 1.
Db	1137 S 1137			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
Query	Match	86.9%	Score 840; DB 11; Length 1774;	DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
Best	Local Similarity	85.6%	Pred. No. 4, 7e-77; Mismatches 11; Indels 0; Gaps 0;	DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
Matches	155; Conservative			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
RESULT	3			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
062001		PRELIMINARY;	PRT; 1774 AA.	DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
ID	062001			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
AC	062001;	PRELIMINARY;	PRT; 1774 AA.	DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
DT	01-NOV-1996 (TREMBLrel. 01, Created)			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
DT	01-OCT-2001 (TREMBLrel. 19, Last annotation update)			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
DE	PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRERECURSOR (XVIII) COLLAGEN			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
DE	(PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
GN	COL18A1.			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
OS	Mus musculus (Mouse)			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
OX	NCBI_TaxID=10090;			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
EN	[1]			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;
RP	SEQUENCE FROM N.A.			DR SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88BF232 CRC64;

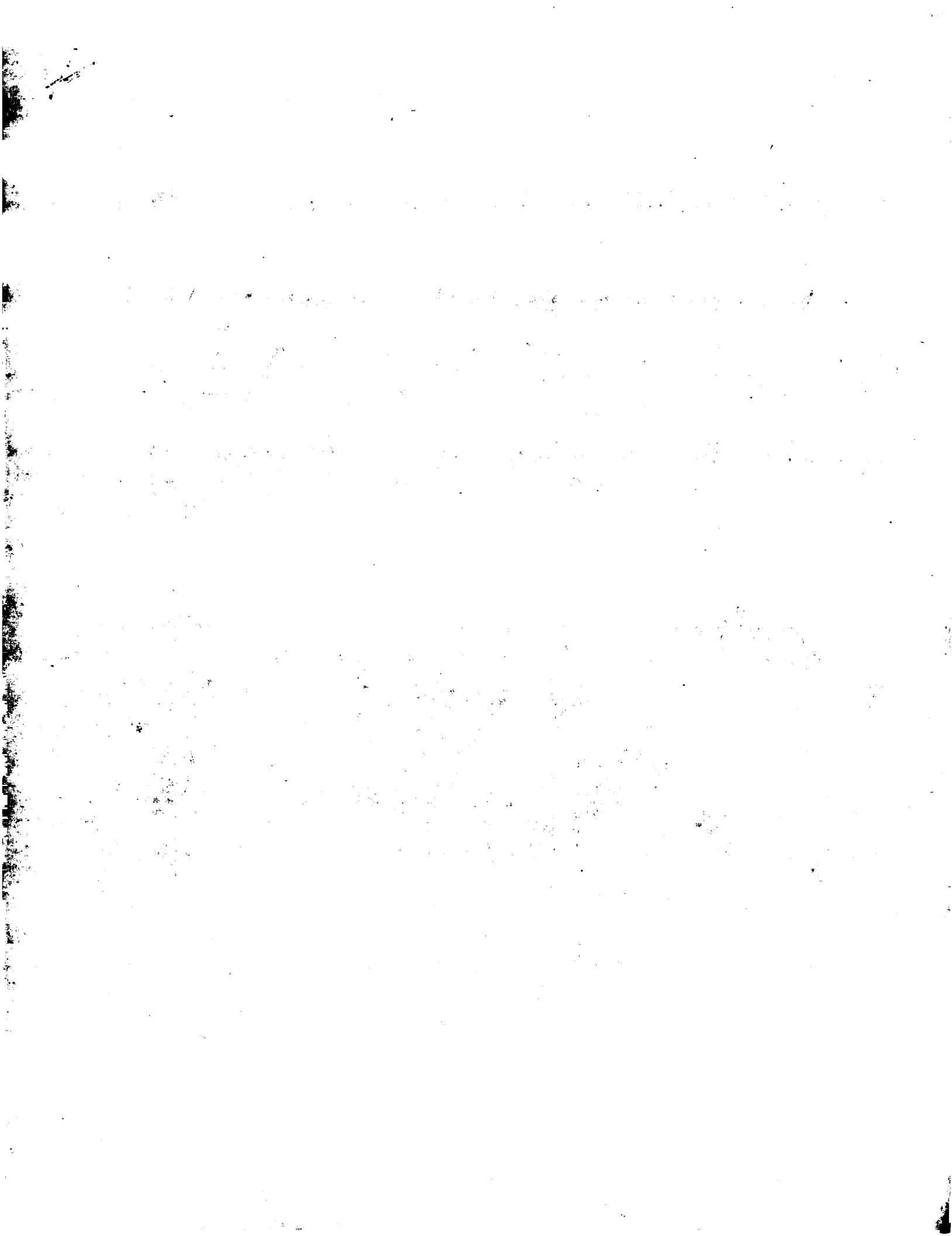
RESULT	4	RT chronic liver injuries.";
QQQZD2	PRELIMINARY;	RT Submitted (FEB-1999) to the EMBL/GenBank/DDBBJ databases.
ID QPOZD2	PRT;	DR EMBL; AR236873; CAB44263.1; -.
AC QQQZD2;		DR HSSP; P39061; 1KOE.
DT 01-MAY-2000 (TREMbrel. 13, Created)		FT NON_TER; 1
DT 01-DEC-2001 (TREMbrel. 19, Last sequence update)		FT NON_TER; 171
DE COLLAGEN XVIII (FRAGMENT).		SQ SEQUENCE 171 AA; 18933 MW; 81B2EE3FC2C88E72 CRC64;
OS rattus norvegicus (Rat).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. OX NCBI_TAXID=10116;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=SPRAGUE-DAWLEY;		
RX MEDLINE=2027226; PubMed=10766159;		
RA Perletti G., Concaro P., Giardini R., Marras E., Piccinini F., Folkman J., Chen L.; "Antitumor activity of endostatin against carcinogen-induced rat primary mammary tumors.", Cancer Res. 60:7793-7796(2000).		
RL DR EMBL; AF189109; ARF00975.1; -.		
DR HSSP; P39061; 1KOE.		
FT NON_TER 1	1	
SEQUENCE 226 AA; 23530 MW; 38B83C0486C0E949 CRC64;		
Query Match 84.8%; Score 820; DB 11; Length 226; Best Local Similarity 84.0%; Pred. No. 3.9e-76; Matches 152; Conservative 14; Mismatches 15; Indels 0; Gaps 0;		
QY 1 HSRRDFQPVHLVALNSPLSGGMRGIRGADFOCFOQARAVGLAGTFRALFLSSRLQDLYSI 60		
Db 43 HTHQDFHPLVHLVALNTPLSGGMRGIRGADFOCFOQARAVGLAGTFRALFLSSRLQDLYSI 102		
QY 61 VRADRAAVPVNLKDELPSWEALFGSGEGPLKPGARIFSDGKVYLHPTWPKSVW 120		
Db 103 VRADRAAVPVNLKDELPSWEALFGSGEGPLKPGARIFSDGKVYLHPTWPKSVW 162		
QY 121 HGSDPNSGRLLTIVSYCEWRTEAPSATQASSLLGRLQKASCHAYIVICIENSFMT 180		
Db 163 HGSDPNSGRLLTIVSYCEWRTEAPSATQASSLLGRLQKASCHAYIVICIENSFMT 222		
QY 181 A 181		
Db 223 S 223		
RESULT 5	6	RT
QPWWWS ID QPWUW5 PRELIMINARY; PRT; 1344 AA.		
AC QPWUW5; PRELIMINARY; PRT; 171 AA.		
DT 01-NOV-1999 (TREMbrel. 12, Created)		
DT 01-NOV-1999 (TREMbrel. 12, Last sequence update)		
DT 01-DEC-2001 (TREMbrel. 19, Last annotation update)		
DE COLLAGEN XVIII PRECURSOR.		
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. OX NCBI_TAXID=9031;		
RN [1]		
RP SEQUENCE FROM N.A.		
QY MEDLINE=98411346; PubMed=9738008; RA Halfter W., Dong S., Schurer B., Cole G.J.; RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.", DE Collagen XVIII, ALPHA (1) CHAIN (FRAGMENT). J. Biol. Chem. 273:25404-25412(1998).		
RN [2]		
RP SEQUENCE FROM N.A.		
RA Halfter W., Dong S.; DR Submitted (Aug-2001) to the EMBL/GenBank/DDBBJ databases.		
DR EMBL; AR083440; ARG3294.2; -.		
DR HSSP; P39061; 1KOE.		
DR InterPro; IPR00087; Collagen.		
DR InterPro; IPR01791; Laminin_G.		
DR InterPro; IPR03129; TSPN.		
DR Pfam; PF01391; Collagen; 7.		
DR Pfam; PF02210; TSPN; 1.		
DR SMART; SM00212; LamG; 1.		
DR SMART; SM00210; TSPN; 1.		
KW Signal.		
FT SIGNAL.		
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;		
Query Match 79.7%; Score 771; DB 13; Length 1344; Best Local Similarity 77.6%; Pred. No. 3.9e-70; Matches 142; Conservative 18; Mismatches 23; Indels 0; Gaps 0;		
QY 1 HSRRDFQPVHLVALNSPLSGGMRGIRGADFOCFOQARAVGLAGTFRALFLSSRLQDLYSI 60		
Db 1161 HVHDFQPVHLVALNTPLSGGMRGIRGADFOCFOQARAVGLAGTFRALFLSSRLQDLYSI 1220		
QY 61 VRADRAAVPVNLKDELPSWEALFGSGEGPLKPGARIFSDGKVYLHPTWPKSVW 120		
"temporospacial expression of collagen XVIII/endostatin in acute and		

QY	121	VRRADPRAFPINLREDEVLFISWNEALFTGSEAPLAGARRAIIISFDGRDQLQDSAWPKWSI 1280	035206	PRELIMINARY;	PRM;	1367 AA.
		HGSDDPGRRLTESYCETWRTEAPSATGQASSLLGGRLIGQSAASCHAYIVLICENSFMT 180	035206			
Db	1281	HGSDAKGRRLPESYCEAWRTDERGTSGQASSLSSGKLLQESACQHAFVVLICENSFMT 1340	035206			
QY	181	ASK 183 [1]	035206			
Db	1341	AAK 1343	035206			
RESULT	7					
O9CRT2						
ID	O9CRT2	PRELIMINARY;	PRM;	160 AA.		
AC	O9CRT2:					
DT	01-JUN-2001 (Tremblel. 17, Created)					
DT	01-JUN-2001 (Tremblel. 17, Last sequence update)					
DE	01-DEC-2001 (Tremblel. 19, Last annotation update)					
DE	PROCOLLAGEN, TYPE XVIII, ALPHA 1 (FRAGMENT).					
GN	COL15A1.					
OS	Mus musculus (Mouse)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=97480713; PubMed=9339358;					
RA	Haag P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,					
RA	Pihlajaniemi T.;					
RT	"Cloning of mouse type XV collagen sequences and mapping of the corresponding gene to 4B1-3. Comparison of mouse and human alpha 1 (XV) collagen sequences indicates divergence in the number of small collagenous domains".					
RT	Genomics 45:31-41(1997);					
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,					
RA	Kadota K., Matsuda H.A., Ashburner M., Baralov S., Casavant T.,					
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,					
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido T., Pesole G., Quackenbush J.,					
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,					
RA	Sakai K., Okido T., Furuno M., Hono H., Baldarelli R., Barsh G.,					
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,					
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Garibaldi M.,					
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,					
RA	Lyonis P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,					
RA	Nordon P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,					
RA	Sasaki H., Sato K., Schoenbach C., Saya T., Shishiba Y., Storch K.-F.,					
RA	Suzuki H., Toyoo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,					
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,					
RT	Hayashizaki Y.; "Functional annotation of a full-length mouse cDNA collection";					
RL	Nature 409:685-690(2001);					
DR	EMBL; AK01292; BAB2949; 1; -.					
DR	HSSP; P39061; 1K0E.					
MDD	MGI:89451; Col18a1.					
FT	NON_TER 1					
SQ	SEQUENCE 160 AA; 17725 MW; 60F853D777C375D2 CRC64;					
RESULT	9					
Q9PQD9	PRELIMINARY;	PRM;	1367 AA.			
ID	Q9PQD9					
AC	Q9PQD9;					
DT	01-MAR-2001 (Tremblel. 16, Created)					
DT	01-MAR-2001 (Tremblel. 16, Last sequence update)					
DE	01-DEC-2001 (Tremblel. 19, Last annotation update)					
DE	TYPE XV COLLAGEN.					
GN	COL15A1.					
OS	Mus musculus (Mouse)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=129/SV;					
RX	MEDLINE=97480713; PubMed=9339358;					
RA	Haag P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,					
RA	Pihlajaniemi T.;					
RT	"Cloning of mouse type XV collagen sequences and mapping of the					

RT	corresponding gene to 4B1-3. Comparison of mouse and human alpha 1	GN	COL15A1.
RT	(XV) collagen sequences indicates divergence in the number of small	OS	Homo sapiens (Human).
RT	collagenous domains.";	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
RL	Genomics 45:31-41(1997).	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN	[2]	OX	NCB_TaxID=9606;
RP	SEQUENCE FROM N.A.	RN	[1]
RC	SRAINE=129/SV;	RP	SEQUENCE FROM N.A.
RX	MEDLINE=20522048; PubMed=11068203;	RX	MEDLINE=9414920; PubMed=8106446;
RA	Eklund L., Muona A., Lietard J., Pihlajaniemi T.;	RA	Kivirikko S., Heinamaki P., Rehn M., Honkanen N., Myers J.C.,
RT	"structure of the mouse type XV collagen gene, Col15a1, comparison	RA	Pihlajaniemi T.;
RT	with the human Col15a1 gene and functional analysis of the promoters	RT	"primary structure of the alpha 1 chain of human type XV collagen and
RT	of both genes.";	RT	exon-intron organization in the 3' region of the corresponding gene.";
RL	Matrix Biol. 19:489-500(2000).	RL	J. Biol. Chem. 269:4773-4779(1994).
DR	EMBL; AF261131; ARG27545.1; -.	RN	[2]
DR	EMBL; AF261139; ARG27545.1; JOINED.	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.
DR	EMBL; AF261110; ARG27545.1; JOINED.	RX	MEDLINE=98316357; PubMed=9651385;
DR	EMBL; AF261111; ARG27545.1; JOINED.	RA	Haug P.M., Muona A., Lietard J., Kivirikko S., Pihlajaniemi T.;
DR	EMBL; AF261112; ARG27545.1; JOINED.	RT	"Complete exon-intron organization of the human gene for the alpha 1
DR	EMBL; AF261113; ARG27545.1; JOINED.	RT	chain of type XV collagen (COL15A1) and comparison with the homologous
DR	EMBL; AF261114; ARG27545.1; JOINED.	RT	COL18A1 gene.";
DR	EMBL; AF261115; ARG27545.1; JOINED.	RT	J. Biol. Chem. 273:1784-17831(1998).
DR	EMBL; AF261116; ARG27545.1; JOINED.	DR	EMBL; L25280; AAC78500.1; -.
DR	EMBL; AF261117; ARG27545.1; JOINED.	DR	EMBL; AR052956; AAC78500.1; JOINED.
DR	EMBL; AF261118; ARG27545.1; JOINED.	DR	EMBL; AR052957; AAC78500.1; JOINED.
DR	EMBL; AF261119; ARG27545.1; JOINED.	DR	EMBL; AR052958; AAC78500.1; JOINED.
DR	EMBL; AF261120; ARG27545.1; JOINED.	DR	EMBL; AR052959; AAC78500.1; JOINED.
DR	EMBL; AF261121; ARG27545.1; JOINED.	DR	EMBL; AR052960; AAC78500.1; JOINED.
DR	EMBL; AF261122; ARG27545.1; JOINED.	DR	EMBL; AR052961; AAC78500.1; JOINED.
DR	EMBL; AF261123; ARG27545.1; JOINED.	DR	EMBL; AR052962; AAC78500.1; JOINED.
DR	EMBL; AF261124; ARG27545.1; JOINED.	DR	EMBL; AR052963; AAC78500.1; JOINED.
DR	EMBL; AF261125; ARG27545.1; JOINED.	DR	EMBL; AR052964; AAC78500.1; JOINED.
DR	EMBL; AF261126; ARG27545.1; JOINED.	DR	EMBL; AR052965; AAC78500.1; JOINED.
DR	EMBL; AF261127; ARG27545.1; JOINED.	DR	EMBL; AR052966; AAC78500.1; JOINED.
DR	EMBL; AF261128; ARG27545.1; JOINED.	DR	EMBL; AR052967; AAC78500.1; JOINED.
DR	EMBL; AF261129; ARG27545.1; JOINED.	DR	EMBL; AR052968; AAC78500.1; JOINED.
DR	EMBL; AF261130; ARG27545.1; JOINED.	DR	EMBL; AR052969; AAC78500.1; JOINED.
DR	EMBL; P39061; 1KOE.	DR	EMBL; AR052970; AAC78500.1; JOINED.
DR	InterPro; IPR000087; Collagen.	DR	EMBL; AR052971; AAC78500.1; JOINED.
DR	InterPro; IPR001791; Laminin_G.	DR	EMBL; AR052972; AAC78500.1; JOINED.
DR	InterPro; IPR003129; TSPN.	DR	EMBL; AR052973; AAC78500.1; JOINED.
DR	Pfam; PF01391; Collagen; 4.	DR	EMBL; AR052974; AAC78500.1; JOINED.
DR	Pfam; PF02210; TSPN; 1.	DR	EMBL; AR052975; AAC78500.1; JOINED.
DR	SMART; SM00282; LamG; 1.	DR	EMBL; L25285; AAC78500.1; JOINED.
DR	SMART; SM00210; TSPN; 1.	DR	EMBL; L25284; AAC78500.1; JOINED.
KW	Collagen.	DR	EMBL; L25283; AAC78500.1; JOINED.
SQ	SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;	DR	EMBL; L25282; AAC78500.1; JOINED.
Query Match	55.9%; Score 541; DB 11; Length 1367;	DR	EMBL; L25281; AAC78500.1; JOINED.
Best Local Similarity	58.2%; Pred. No. 1.6e-45;	DR	HSSP; P39061; 1KOE.
Matches	103; Conservative 28; Mismatches 42; Indels 4; Gaps 1;	DR	InterPro; IPR000087; Collagen.
Qy	7 QPVHLVALNSLSSGGGIRGADFOCQFQQRAGVLAGTFRALFLSLRQLDLYSIVRRADR 66	DR	InterPro; IPR001791; Laminin_G.
Db	1194 RPVLHLVALNTPVAGDIR---ADFOCFOQARAGLSTFRALFLSHLQLDSLTVVRAER 1249	DR	InterPro; IPR003129; TSPN.
Qy	67 AAVPIVNLKDELLPSMEALFGSGMGGIRGADFOCQFQQRAGVLAGTFRALFLSLRQLDLYSIVRRADR 66	DR	Pfam; PF01391; Collagen; 4.
Db	1250 FGLPIVNLKQVLFNNMDISIFSGDGSOENTHPPYIYSDGRDQMTDPSWPKVWHGSNPY 1309	DR	Pfam; PF02210; TSPN; 1.
RESULT	10	DR	SMART; SM00282; LamG; 1.
QY4W4	PRELIMINARY; PRT; 1388 AA.	DR	SMART; SM00202; Laminin_G.
ID	QY4W4	DR	EMBL; L25281; AAC78500.1; JOINED.
AC	QY4W4; 01-NOV-1999 (TREMBrel. 12, Created)	DR	HSSP; P39061; 1KOE.
DT	01-NOV-1999 (TREMBrel. 12, Last sequence update)	DR	InterPro; IPR000087; Collagen.
DT	01-DEC-2001 (TREMBrel. 19, Last annotation update)	DR	InterPro; IPR001791; Laminin_G.
DE	TYPE XV COLLAGEN.	DR	InterPro; IPR003129; TSPN.
Query Match	55.9%; Score 530; DB 4; Length 1388;	DR	InterPro; IPR000087; Collagen.
Best Local Similarity	55.9%; Pred. No. 2.2e-45;	DR	InterPro; IPR001791; Laminin_G.
Matches	99; Conservative 31; Mismatches 43; Indels 4; Gaps 1;	DR	InterPro; IPR003129; TSPN.
Qy	7 QPVHLVALNSLSSGGGIRGADFOCQFQQRAGVLAGTFRALFLSLRQLDLYSIVRRADR 66	DR	Pfam; PF01391; Collagen; 4.
Db	1215 KPAHLHLANLMPVSGDIR---ADFOCFOQARAGLSTFRALFLSHLQLDSLTVVRAER 1270	DR	SMART; SM00210; AAC78500.1; JOINED.
Qy	67 AAVPIVNLKDELLPSMEALFGSGMGGIRGADFOCQFQQRAGVLAGTFRALFLSLRQLDLYSIVRRADR 126	DR	HSSP; P39061; 1KOE.
Db	1271 YSLPIVNLKQVLFNNMDISIFSGDGSOENTHPPYIYSDGRDQMTDPSWPKVWHGSNPY 1330	DR	InterPro; IPR000087; Collagen.
Qy	127 GRRLTESCGCTWRTEAPSATGQASSLIGGLGQSAASCHHAYIVCIENSMTASK 183	DR	InterPro; IPR001791; Laminin_G.
Db	1310 GVRLNDKYCEAWRTDMAVTGFPASPLSTGKLDQKAYSCANRLLIVCIENSFWIDTR 1366	DR	InterPro; IPR003129; TSPN.

Db	1331	GVRLVDNYCEAWRTDAVIGLASPLSTGKILDQKAYSCANRIVLICIENSFTMDAR	1387
RESULT	11		
ID	Q96T70	PRELIMINARY;	PRT; 102 AA.
AC	Q96T70;		
DT	01-DEC-2001 (TREMBrel. 19, Created)		
DT	01-DEC-2001 (TREMBrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBrel. 19, Last annotation update)		
DE	ENDOSTATIN VARIANT (FRAGMENT).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TAXID=9606;		
RN	{1}		
SEQUENCE	FROM N A.		
RA	"Endostatin promotes delayed secondary damage following traumatic brain injury." Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.		
RA	Deininger M.H., Trautmann K., Schluessener H.J., RT		
RA	Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL: AF333247; AAK50626.1; -.		
FT	NON_TER 1		
FT	NON_TER 102 102		
SEQUENCE	102 AA; 11147 MW; ECAC47AA6420947D CRC64;		
Query Match	54.7%; Score 529; DB 4; Length 102;		
Best Local Similarity	98.0%; Pred. No. 9, 1e-47;		
Matches	100; Conservative 0; Mismatches 2; Indels 0; Gaps 0; O;		
Qy	71 IVNLKDELLFPSWEALFSGSGPPLKGARIFSFQDGKVLRHPTWPKQSKVWHGSDPNRRRL 130		
Db	1 IVNLKDELLFPSWEALFSGSGPPLKGARIFSFQDGKVLRHPTWPKQSKVWHGSDPNRRRL 60		
Qy	131 TESYCETWRTEAPSATGQASSLLGGRLGQSAASSHAYIVL 172		
Db	61 TESYCETWRTEAPSATGQASSLLGGRLGQSAASSHAYIVL 102		
RESULT	12		
Q9VS09	PRELIMINARY; PRT; 581 AA.		
AC	Q9VS09;		
DT	01-MAY-2000 (TREMBrel. 13, Created)		
DT	01-MAY-2000 (TREMBrel. 13, Last sequence update)		
DT	01-JUN-2001 (TREMBrel. 17, Last annotation update)		
DE	CG8645 PROTEIN.		
GN	Drosophila melanogaster (Fruit fly).		
OS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyoidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanades P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Winkler M.D., Zhang Q., Chen L.X., Sutton G.G., Wortman J.R., Yandell S., Ashburner M., Henderson S.N., Brandon R.C., Rogers Y.-H.C., Blakely R.G., Champ M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Heft G., Nelson C.R., Miklos G.L.G., Abril J.F., Agayam A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Bhandari D., Bolshakov S., Borckson D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Centee A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Paoli A., Delcher A., Deng Z., Mayes A.D., Dew I., Dietz S.M., Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin R.J., Evangelista C.C., Ferraz C., Ferreira S., Fieischmann W.,		
SO	SEQUENCE	581 AA; 60772 MW; 19EC1E48CB477EB7 CRC64;	
Query Match	44.5%; Score 430.5; DB 5; Length 581;		
Best Local Similarity	48.8%; Pred. No. 1, 2e-35;		
Matches	82; Conservative 29; Mismatches 56; Indels 1; Gaps 1; O;		
Db	373 PTLRVAALNEPSTGDLQIGRAGDFACYQRQRARRAGLGTKEFLSSRVQNLDITVPRAD- 431		
Qy	68 AVPIVNLKDELLFPSWEALFSGSGPPLKGARIFSFQDGKVLRHPTWPKQSKVWHGSDPN 127		
Db	432 DLPPVNTRGVLFLNSKQGIGNGQGPFSQAPRISYSFKVNMIDSTWPMMWVHGSPLNG 491		
Qy	128 RRLTESYCETWRTEAPSATGQASSLLGGRLGQSAASSHAYIVL 175		
Db	492 ERSMDTYCDAWHSGDHLKGSPFASNLQDHKLIEQKROSCDSKLILC 539		
RESULT	13		
ID	Q17866	PRELIMINARY;	PRT; 650 AA.
AC	Q17866;		
DT	01-JAN-1998 (TREMBrel. 05, Created)		
DT	01-AUG-1998 (TREMBrel. 07, Last sequence update)		
DT	01-DEC-2001 (TREMBrel. 19, Last annotation update)		
DE	CIE-IC PROTEIN.		
GN	Caenorhabditis elegans.		
OS	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peletorinae; Caenorhabditis.		
OC	NCBI_TAXID=6239;		
RN	[1]		
SEQUENCE	FROM N A.		
RA	MEDLINE=21157401; PubMed=11257122;		
RA	Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J., Kramer J.M., "The N1/endostatin domain of <i>Caenorhabditis elegans</i> type XVIII collagen affects cell migration and axon guidance.," J. Cell Biol. 152:1219-1232(2001).		
DR	EMBL: AF164959; AADM7825.1; -.		
DR	HSSP; P39061; IROE.		





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:12:33 ; Search time 11.91 Seconds

(without alignments)

594.935 Million cell updates/sec

Title: DS-10-080-797-1

Perfect score: 967

Sequence: HSHRDFQPVHLVALNSPLS.....SCHHAYIWLICIENSFMFTASK 183

Scoring table:

BLOSUM62, Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%, Maximum Match 100%

Listing first 45 summaries

Database : Swissprot\_40;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	967	100.0	1516	1 CAIH_HUMAN	P39060 homo sapien
2	840	86.9	1527	1 CAIH_MOUSE	P39061 mus musculus
3	529	54.7	1388	1 CAIE_HUMAN	P39059 homo sapien
4	76	7.9	556	1 GLI_GICK	P55878 gallus gallus
5	75	7.8	271	1 Y4BG_RHISN	P55374 rhizobium sphaericum
6	74	7.7	296	1 CRK_XENLA	P87378 xenopus laevis
7	74	7.7	780	1 STRN_RAT	P70483 rattus norvegicus
8	74	7.7	1215	1 HDAG_HUMAN	Opubn7 homo sapien
9	73	7.5	512	1 PPX_ECOLI	P29014 escherichia coli
10	72	7.5	884	1 P87_HYCTU	P05654 mycobacterium tuberculosis
11	72	7.4	304	1 CRK_MOUSE	P46108 homo sapiens
12	72	7.4	304	1 CRK_RAT	064010 mus musculus
13	72	7.4	304	1 CRK_MOUSE	063768 rattus norvegicus
14	72	7.4	613	1 HSP75_CANAL	P87222 candida albicans
15	70.5	7.3	312	1 NRTH_CHICK	P55806 gallus gallus
16	70	7.2	780	1 STRN_MOUSE	P55100 mus musculus
17	69	7.1	266	1 CB2L_SINAL	P13851 sinapis alba
18	69	7.1	309	1 YHCC_ECOLI	P45476 escherichia coli
19	69	7.1	326	1 TMEF_SEME	Q00460 pseudomonas aeruginosa
20	69	7.1	364	1 YM2B_MYCUT	Q010512 mycobacterium tuberculosis
21	69	7.1	462	1 MYCN_MOUSE	P03966 mus musculus
22	69	7.1	1233	1 NHE3_HUMAN	Q14957 homo sapiens
23	68.5	7.1	390	1 PGK_BUCAT	P57525 buchnera aphidis
24	68.5	7.1	1289	1 C5AB_BACUD	Q45753 bacillus thuringiensis
25	68	7.0	953	1 CRA_HUMAN	Q9y239 homo sapiens
26	68	7.0	999	1 MERK_HUMAN	Q12866 homo sapiens
27	68	7.0	3491	1 ERYL_SADER	Q03131 saccharopolysaccharide
28	67.5	7.0	145	1 ANF_RANCA	P18890 ranunculus
29	67.5	7.0	317	1 GGH_RAT	Q02867 rattus norvegicus
30	67.5	7.0	536	1 CAR9_HUMAN	Q02867 rattus norvegicus
31	67	7.0	579	1 FAAH_MOUSE	Q08914 mus musculus
32	67.5	7.0	594	1 MGII_MOUSE	Q67010 mus musculus
33	7.0			1 PRPE_ECOLI	P77495 escherichia coli

#### ALIGNMENTS

RESULT 1	CAIH_HUMAN	STANDARD; PRT; 1516 AA.
ID CAIH_HUMAN	SEQUENCE FROM N.A.	[2]
RA MEDLINE-98164096; PubMed=10830953;	RX	
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E., Ohki M., Takayama T., Sakai Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillihabel M., Schudy A., Zimmermann W., Rosenthal A., Kudo J., Shibuya K., Kawasaki K., Asakawa S., Anttonenakis S.E., Shintani A., Sasaki T., Nakamizo K., Mitsuyama S., Anttonenakis S.E., Minoshima S., Shimizu N., Nordstek G., Hornischer K., Brandt P., Schafre M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H., Ramsier J., Beck A., Klages S., Hennig S., Riessemann L., Daidone E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.; "The DNA sequence of human chromosome 21.", RT Nature 405:311-319(2000).	RL	
RA [3]	RN	
RA SEQUENCE OF 834-1516 FROM N.A.	RX	
RA MEDLINE-94245237; PubMed=8888291; Oh S.P., Warmann M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S., RA Olsen B.R.; "Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the alpha 1(XVIII) collagen gene to mouse chromosome 10 and human chromosome 21.", Genomics 19:494-499(1994).	RT	
RA [4]	RN	
RA SEQUENCE OF 1334-1516 FROM N.A.	RT	
RA TISSUE=placenta; Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.; "Cloning and expression of human endostatin gene in Escherichia coli.", Submitted (SPP-1999) to the EMBL/Genbank/DBJ databases.	RT	
RA [5]	RL	
RA INVOLVEMENT IN KNOBLOCH SYNDROME.	RN	



RESULT 2  
 CATH\_MOUSE STANDARD; PRN; 1527 AA.  
 ID CAIH\_MOUSE STANDARD;  
 AC P39061; Q62002; Q61437;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(XVIII) chain precursor [Contains:  
 GN COL18A1.  
 OS MUS musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NCBI\_TaxID=10090;  
 RP [1] SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC STRAIN-BALB/C; TISSUE-Liver;  
 RX MEDLINE=94245707; PubMed=8188673;  
 RA Rehn M.V., Hintikka E., Pihalajaniemi T.;  
 RT "Primary structure of the mouse gene for the alpha-1 chain of type XVIII collagen (COL18A1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters.",  
 RT Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.  
 RN [3] SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).  
 RX MEDLINE=94240112; PubMed=8188394;  
 RA Rein M.V., Pihalajaniemi T.;  
 RT "Alpha 1 (XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen.",  
 RT Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).  
 RN [4] SEQUENCE OF 240-1527 FROM N.A.  
 RC TISSUE-liver;  
 RX MEDLINE=94240111; PubMed=8188393;  
 RA Oh S.P., Kamagata Y., Murakami Y., Timmons S., Coshima A., Olsen B.R.;  
 RT "Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins.",  
 RT Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).  
 RN [5] RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.  
 RX MEDLINE=97160848; PubMed=9008168;  
 RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vassios G., Lane W.S.,  
 RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;  
 RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor growth",  
 RT Cell 88: 277-285(1997).  
 RN [6] RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.  
 RX MEDLINE=98169382; PubMed=9301087;  
 RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;  
 RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A resolution.",  
 RL EMBJ 17:1656-1664(1998).  
 CC -!- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH FACTOR SIGNALING.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOMERS; A SHORT FORM AND A LONG FORM (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPlicing.  
 CC -!- PTM PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-T) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH

CC .-. INTERRUPTED HELICES (FACIT) FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions. There is no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [license@isb-sib.ch](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch)).  
 CC  
 CC DR EMBL; L16898; AAA37434.1; -.  
 DR EMBL; U03714; AAC2901.1; -.  
 DR EMBL; U34606; AAC2901.1; JOINED.  
 DR EMBL; U34608; AAC2901.1; JOINED.  
 DR EMBL; U34610; AAC2901.1; JOINED.  
 DR EMBL; U34611; AAC2901.1; JOINED.  
 DR EMBL; U34612; AAC2901.1; JOINED.  
 DR EMBL; U34613; AAC2901.1; JOINED.  
 DR EMBL; U34614; AAC2901.1; JOINED.  
 DR EMBL; U34615; AAC2901.1; JOINED.  
 DR EMBL; U34616; AAC2901.1; JOINED.  
 DR EMBL; U34617; AAC2901.1; JOINED.  
 DR EMBL; U34618; AAC2901.1; JOINED.  
 DR EMBL; U34619; AAC2901.1; JOINED.  
 DR EMBL; U34620; AAC2901.1; JOINED.  
 DR EMBL; U34621; AAC2902.1; JOINED.  
 DR EMBL; U34622; AAC2902.1; JOINED.  
 DR EMBL; U34623; AAC2902.1; JOINED.  
 DR EMBL; U34624; AAC2902.1; JOINED.  
 DR EMBL; U34625; AAC2902.1; JOINED.  
 DR EMBL; U34626; AAC2902.1; JOINED.  
 DR EMBL; U34627; AAC2902.1; JOINED.  
 DR EMBL; U34628; AAC2902.1; JOINED.  
 DR EMBL; U34629; AAC2902.1; JOINED.  
 DR EMBL; U34630; AAC2902.1; JOINED.  
 DR EMBL; U34631; AAC2902.1; JOINED.  
 DR EMBL; U34632; AAC2902.1; JOINED.  
 DR EMBL; U34633; AAC2902.1; JOINED.  
 DR EMBL; U34634; AAC2902.1; JOINED.  
 DR EMBL; U34635; AAC2902.1; JOINED.  
 DR EMBL; U34636; AAC2902.1; JOINED.  
 DR EMBL; U34637; AAC2902.1; JOINED.  
 DR EMBL; U34638; AAC2902.1; JOINED.  
 DR EMBL; U34639; AAC2902.1; JOINED.  
 DR EMBL; U34640; AAC2902.1; JOINED.  
 DR EMBL; U34641; AAC2902.1; JOINED.  
 DR EMBL; U34642; AAC2902.1; JOINED.  
 DR EMBL; U34643; AAC2902.1; JOINED.  
 DR EMBL; U34644; AAC2902.1; JOINED.  
 DR EMBL; U34645; AAC2902.1; JOINED.  
 DR EMBL; U34646; AAC2902.1; JOINED.  
 DR EMBL; U34647; AAC2902.1; JOINED.  
 DR EMBL; U34648; AAC2902.1; JOINED.  
 DR EMBL; U34649; AAC2902.1; JOINED.  
 DR EMBL; U34650; AAC2902.1; JOINED.  
 DR EMBL; U34651; AAC2902.1; JOINED.  
 DR EMBL; U34652; AAC2902.1; JOINED.  
 DR EMBL; U34653; AAC2902.1; JOINED.  
 DR EMBL; U34654; AAC2902.1; JOINED.  
 DR EMBL; U34655; AAC2902.1; JOINED.  
 DR EMBL; U34656; AAC2902.1; JOINED.  
 DR EMBL; U34657; AAC2902.1; JOINED.  
 DR EMBL; U34658; AAC2902.1; JOINED.  
 DR EMBL; U34659; AAC2902.1; JOINED.  
 DR EMBL; U34660; AAC2902.1; JOINED.  
 DR EMBL; U34661; AAC2902.1; JOINED.  
 DR EMBL; U34662; AAC2902.1; JOINED.  
 DR EMBL; U34663; AAC2902.1; JOINED.  
 DR EMBL; U34664; AAC2902.1; JOINED.  
 DR EMBL; U34665; AAC2902.1; JOINED.  
 DR EMBL; U34666; AAC2902.1; JOINED.  
 DR EMBL; U34667; AAC2902.1; JOINED.  
 DR EMBL; U34668; AAC2902.1; JOINED.  
 DR EMBL; U34669; AAC2902.1; JOINED.  
 DR EMBL; U34670; AAC2902.1; JOINED.  
 DR EMBL; U34671; AAC2902.1; JOINED.  
 DR EMBL; U34672; AAC2902.1; JOINED.  
 DR EMBL; U34673; AAC2902.1; JOINED.  
 DR EMBL; U34674; AAC2902.1; JOINED.  
 DR EMBL; U34675; AAC2902.1; JOINED.  
 DR EMBL; U34676; AAC2902.1; JOINED.  
 DR EMBL; U34677; AAC2902.1; JOINED.  
 DR EMBL; U34678; AAC2902.1; JOINED.  
 DR EMBL; U34679; AAC2902.1; JOINED.  
 DR EMBL; U34680; AAC2902.1; JOINED.  
 DR EMBL; U34681; AAC2902.1; JOINED.  
 DR EMBL; U34682; AAC2902.1; JOINED.  
 DR EMBL; U34683; AAC2902.1; JOINED.  
 DR EMBL; U34684; AAC2902.1; JOINED.  
 DR EMBL; U34685; AAC2902.1; JOINED.  
 DR EMBL; U34686; AAC2902.1; JOINED.  
 DR EMBL; U34687; AAC2902.1; JOINED.  
 DR EMBL; U34688; AAC2902.1; JOINED.  
 DR EMBL; U34689; AAC2902.1; JOINED.  
 DR EMBL; U34690; AAC2902.1; JOINED.  
 DR EMBL; U34691; AAC2902.1; JOINED.  
 DR EMBL; U34692; AAC2902.1; JOINED.  
 DR EMBL; U34693; AAC2902.1; JOINED.  
 DR EMBL; U34694; AAC2902.1; JOINED.  
 DR EMBL; U34695; AAC2902.1; JOINED.  
 DR EMBL; U34696; AAC2902.1; JOINED.  
 DR EMBL; U34697; AAC2902.1; JOINED.  
 DR EMBL; U34698; AAC2902.1; JOINED.  
 DR EMBL; U34699; AAC2902.1; JOINED.  
 DR EMBL; U34700; AAC2902.1; JOINED.  
 DR EMBL; U34701; AAC2902.1; JOINED.  
 DR EMBL; U34702; AAC2902.1; JOINED.  
 DR EMBL; U34703; AAC2902.1; JOINED.  
 DR EMBL; U34704; AAC2902.1; JOINED.  
 DR EMBL; U34705; AAC2902.1; JOINED.  
 DR EMBL; U34706; AAC2902.1; JOINED.  
 DR EMBL; U34707; AAC2902.1; JOINED.  
 DR EMBL; U34708; AAC2902.1; JOINED.  
 DR EMBL; U34709; AAC2902.1; JOINED.  
 DR EMBL; U34710; AAC2902.1; JOINED.  
 DR EMBL; U34711; AAC2902.1; JOINED.  
 DR EMBL; U34712; AAC2902.1; JOINED.  
 DR EMBL; U34713; AAC2902.1; JOINED.  
 DR EMBL; U34714; AAC2902.1; JOINED.  
 DR EMBL; U34715; AAC2902.1; JOINED.  
 DR EMBL; U34716; AAC2902.1; JOINED.  
 DR EMBL; U34717; AAC2902.1; JOINED.  
 DR EMBL; U34718; AAC2902.1; JOINED.  
 DR EMBL; U34719; AAC2902.1; JOINED.  
 DR EMBL; U34720; AAC2902.1; JOINED.  
 DR EMBL; U34721; AAC2902.1; JOINED.  
 DR EMBL; U34722; AAC2902.1; JOINED.  
 DR EMBL; U34723; AAC2902.1; JOINED.  
 DR EMBL; U34724; AAC2902.1; JOINED.  
 DR EMBL; U34725; AAC2902.1; JOINED.  
 DR EMBL; U34726; AAC2902.1; JOINED.  
 DR EMBL; U34727; AAC2902.1; JOINED.  
 DR EMBL; U34728; AAC2902.1; JOINED.  
 DR EMBL; U34729; AAC2902.1; JOINED.  
 DR EMBL; U34730; AAC2902.1; JOINED.  
 DR EMBL; U34731; AAC2902.1; JOINED.  
 DR EMBL; U34732; AAC2902.1; JOINED.  
 DR EMBL; U34733; AAC2902.1; JOINED.  
 DR EMBL; U34734; AAC2902.1; JOINED.  
 DR EMBL; U34735; AAC2902.1; JOINED.  
 DR EMBL; U34736; AAC2902.1; JOINED.  
 DR EMBL; U34737; AAC2902.1; JOINED.  
 DR EMBL; U34738; AAC2902.1; JOINED.  
 DR EMBL; U34739; AAC2902.1; JOINED.  
 DR EMBL; U34740; AAC2902.1; JOINED.  
 DR EMBL; U34741; AAC2902.1; JOINED.  
 DR EMBL; U34742; AAC2902.1; JOINED.  
 DR EMBL; U34743; AAC2902.1; JOINED.  
 DR EMBL; U34744; AAC2902.1; JOINED.  
 DR EMBL; U34745; AAC2902.1; JOINED.  
 DR EMBL; U34746; AAC2902.1; JOINED.  
 DR EMBL; U34747; AAC2902.1; JOINED.  
 DR EMBL; U34748; AAC2902.1; JOINED.  
 DR EMBL; U34749; AAC2902.1; JOINED.  
 DR EMBL; U34750; AAC2902.1; JOINED.  
 DR EMBL; U34751; AAC2902.1; JOINED.  
 DR EMBL; U34752; AAC2902.1; JOINED.  
 DR EMBL; U34753; AAC2902.1; JOINED.  
 DR EMBL; U34754; AAC2902.1; JOINED.  
 DR EMBL; U34755; AAC2902.1; JOINED.  
 DR EMBL; U34756; AAC2902.1; JOINED.  
 DR EMBL; U34757; AAC2902.1; JOINED.  
 DR EMBL; U34758; AAC2902.1; JOINED.  
 DR EMBL; U34759; AAC2902.1; JOINED.  
 DR EMBL; U34760; AAC2902.1; JOINED.  
 DR EMBL; U34761; AAC2902.1; JOINED.  
 DR EMBL; U34762; AAC2902.1; JOINED.  
 DR EMBL; U34763; AAC2902.1; JOINED.  
 DR EMBL; U34764; AAC2902.1; JOINED.  
 DR EMBL; U34765; AAC2902.1; JOINED.  
 DR EMBL; U34766; AAC2902.1; JOINED.  
 DR EMBL; U34767; AAC2902.1; JOINED.  
 DR EMBL; U34768; AAC2902.1; JOINED.  
 DR EMBL; U34769; AAC2902.1; JOINED.  
 DR EMBL; U34770; AAC2902.1; JOINED.  
 DR EMBL; U34771; AAC2902.1; JOINED.  
 DR EMBL; U34772; AAC2902.1; JOINED.  
 DR EMBL; U34773; AAC2902.1; JOINED.  
 DR EMBL; U34774; AAC2902.1; JOINED.  
 DR EMBL; U34775; AAC2902.1; JOINED.  
 DR EMBL; U34776; AAC2902.1; JOINED.  
 DR EMBL; U34777; AAC2902.1; JOINED.  
 DR EMBL; U34778; AAC2902.1; JOINED.  
 DR EMBL; U34779; AAC2902.1; JOINED.  
 DR EMBL; U34780; AAC2902.1; JOINED.  
 DR EMBL; U34781; AAC2902.1; JOINED.  
 DR EMBL; U34782; AAC2902.1; JOINED.  
 DR EMBL; U34783; AAC2902.1; JOINED.  
 DR EMBL; U34784; AAC2902.1; JOINED.  
 DR EMBL; U34785; AAC2902.1; JOINED.  
 DR EMBL; U34786; AAC2902.1; JOINED.  
 DR EMBL; U34787; AAC2902.1; JOINED.  
 DR EMBL; U34788; AAC2902.1; JOINED.  
 DR EMBL; U34789; AAC2902.1; JOINED.  
 DR EMBL; U34790; AAC2902.1; JOINED.  
 DR EMBL; U34791; AAC2902.1; JOINED.  
 DR EMBL; U34792; AAC2902.1; JOINED.  
 DR EMBL; U34793; AAC2902.1; JOINED.  
 DR EMBL; U34794; AAC2902.1; JOINED.  
 DR EMBL; U34795; AAC2902.1; JOINED.  
 DR EMBL; U34796; AAC2902.1; JOINED.  
 DR EMBL; U34797; AAC2902.1; JOINED.  
 DR EMBL; U34798; AAC2902.1; JOINED.  
 DR EMBL; U34799; AAC2902.1; JOINED.  
 DR EMBL; U34800; AAC2902.1; JOINED.  
 DR EMBL; U34801; AAC2902.1; JOINED.  
 DR EMBL; U34802; AAC2902.1; JOINED.  
 DR EMBL; U34803; AAC2902.1; JOINED.  
 DR EMBL; U34804; AAC2902.1; JOINED.  
 DR EMBL; U34805; AAC2902.1; JOINED.  
 DR EMBL; U34806; AAC2902.1; JOINED.  
 DR EMBL; U34807; AAC2902.1; JOINED.  
 DR EMBL; U34808; AAC2902.1; JOINED.  
 DR EMBL; U34809; AAC2902.1; JOINED.  
 DR EMBL; U34810; AAC2902.1; JOINED.  
 DR EMBL; U34811; AAC2902.1; JOINED.  
 DR EMBL; U34812; AAC2902.1; JOINED.  
 DR EMBL; U34813; AAC2902.1; JOINED.  
 DR EMBL; U34814; AAC2902.1; JOINED.  
 DR EMBL; U34815; AAC2902.1; JOINED.  
 DR EMBL; U34816; AAC2902.1; JOINED.  
 DR EMBL; U34817; AAC2902.1; JOINED.  
 DR EMBL; U34818; AAC2902.1; JOINED.  
 DR EMBL; U34819; AAC2902.1; JOINED.  
 DR EMBL; U34820; AAC2902.1; JOINED.  
 DR EMBL; U34821; AAC2902.1; JOINED.  
 DR EMBL; U34822; AAC2902.1; JOINED.  
 DR EMBL; U34823; AAC2902.1; JOINED.  
 DR EMBL; U34824; AAC2902.1; JOINED.  
 DR EMBL; U34825; AAC2902.1; JOINED.  
 DR EMBL; U34826; AAC2902.1; JOINED.  
 DR EMBL; U34827; AAC2902.1; JOINED.  
 DR EMBL; U34828; AAC2902.1; JOINED.  
 DR EMBL; U34829; AAC2902.1; JOINED.  
 DR EMBL; U34830; AAC2902.1; JOINED.  
 DR EMBL; U34831; AAC2902.1; JOINED.  
 DR EMBL; U34832; AAC2902.1; JOINED.  
 DR EMBL; U34833; AAC2902.1; JOINED.  
 DR EMBL; U34834; AAC2902.1; JOINED.  
 DR EMBL; U34835; AAC2902.1; JOINED.  
 DR EMBL; U34836; AAC2902.1; JOINED.  
 DR EMBL; U34837; AAC2902.1; JOINED.  
 DR EMBL; U34838; AAC2902.1; JOINED.  
 DR EMBL; U34839; AAC2902.1; JOINED.  
 DR EMBL; U34840; AAC2902.1; JOINED.  
 DR EMBL; U34841; AAC2902.1; JOINED.  
 DR EMBL; U34842; AAC2902.1; JOINED.  
 DR EMBL; U34843; AAC2902.1; JOINED.  
 DR EMBL; U34844; AAC2902.1; JOINED.  
 DR EMBL; U34845; AAC2902.1; JOINED.  
 DR EMBL; U34846; AAC2902.1; JOINED.  
 DR EMBL; U34847; AAC2902.1; JOINED.  
 DR EMBL; U34848; AAC2902.1; JOINED.  
 DR EMBL; U34849; AAC2902.1; JOINED.  
 DR EMBL; U34850; AAC2902.1; JOINED.  
 DR EMBL; U34851; AAC2902.1; JOINED.  
 DR EMBL; U34852; AAC2902.1; JOINED.  
 DR EMBL; U34853; AAC2902.1; JOINED.  
 DR EMBL; U34854; AAC2902.1; JOINED.  
 DR EMBL; U34855; AAC2902.1; JOINED.  
 DR EMBL; U34856; AAC2902.1; JOINED.  
 DR EMBL; U34857; AAC2902.1; JOINED.  
 DR EMBL; U34858; AAC2902.1; JOINED.  
 DR EMBL; U34859; AAC2902.1; JOINED.  
 DR EMBL; U34860; AAC2902.1; JOINED.  
 DR EMBL; U34861; AAC2902.1; JOINED.  
 DR EMBL; U34862; AAC2902.1; JOINED.  
 DR EMBL; U34863; AAC2902.1; JOINED.  
 DR EMBL; U34864; AAC2902.1; JOINED.  
 DR EMBL; U34865; AAC2902.1; JOINED.  
 DR EMBL; U34866; AAC2902.1; JOINED.  
 DR EMBL; U34867; AAC2902.1; JOINED.  
 DR EMBL; U34868; AAC2902.1; JOINED.  
 DR EMBL; U34869; AAC2902.1; JOINED.  
 DR EMBL; U34870; AAC2902.1; JOINED.  
 DR EMBL; U34871; AAC2902.1; JOINED.  
 DR EMBL; U34872; AAC2902.1; JOINED.  
 DR EMBL; U34873; AAC2902.1; JOINED.  
 DR EMBL; U34874; AAC2902.1; JOINED.  
 DR EMBL; U34875; AAC2902.1; JOINED.  
 DR EMBL; U34876; AAC2902.1; JOINED.  
 DR EMBL; U34877; AAC2902.1; JOINED.  
 DR EMBL; U34878; AAC2902.1; JOINED.  
 DR EMBL; U34879; AAC2902.1; JOINED.  
 DR EMBL; U34880; AAC2902.1; JOINED.  
 DR EMBL; U34881; AAC2902.1; JOINED.  
 DR EMBL; U34882; AAC2902.1; JOINED.  
 DR EMBL; U34883; AAC2902.1; JOINED.  
 DR EMBL; U34884; AAC2902.1; JOINED.  
 DR EMBL; U34885; AAC2902.1; JOINED.  
 DR EMBL; U34886; AAC2902.1; JOINED.  
 DR EMBL; U34887; AAC2902.1; JOINED.  
 DR EMBL; U34888; AAC2902.1; JOINED.  
 DR EMBL; U34889; AAC2902.1; JOINED.  
 DR EMBL; U34890; AAC2902.1; JOINED.  
 DR EMBL; U34891; AAC2902.1; JOINED.  
 DR EMBL; U34892; AAC2902.1; JOINED.  
 DR EMBL; U34893; AAC2902.1; JOINED.  
 DR EMBL; U34894; AAC2902.1; JOINED.  
 DR EMBL; U34895; AAC2902.1; JOINED.  
 DR EMBL; U34896; AAC2902.1; JOINED.  
 DR EMBL; U34897; AAC2902.1; JOINED.  
 DR EMBL; U34898; AAC2902.1; JOINED.  
 DR EMBL; U34899; AAC2902.1; JOINED.  
 DR EMBL; U34900; AAC2902.1; JOINED.  
 DR EMBL; U34901; AAC2902.1; JOINED.  
 DR EMBL; U34902; AAC2902.1; JOINED.  
 DR EMBL; U34903; AAC2902.1; JOINED.  
 DR EMBL; U34904; AAC2902.1; JOINED.  
 DR EMBL; U34905; AAC2902.1; JOINED.  
 DR EMBL; U34906; AAC2902.1; JOINED.  
 DR EMBL; U34907; AAC2902.1; JOINED.  
 DR EMBL; U34908; AAC2902.1; JOINED.  
 DR EMBL; U34909; AAC2902.1; JOINED.  
 DR EMBL; U34910; AAC2902.1; JOINED.  
 DR EMBL; U34911; AAC2902.1; JOINED.  
 DR EMBL; U34912; AAC2902.1; JOINED.  
 DR EMBL; U34913; AAC2902.1; JOINED.  
 DR EMBL; U34914; AAC2902.1; JOINED.  
 DR EMBL; U34915; AAC2902.1; JOINED.  
 DR EMBL; U34916; AAC2902.1; JOINED.  
 DR EMBL; U34917; AAC2902.1; JOINED.  
 DR EMBL; U34918; AAC2902.1; JOINED.  
 DR EMBL; U34919; AAC2902.1; JOINED.  
 DR EMBL; U34920; AAC2902.1; JOINED.  
 DR EMBL; U34921; AAC2902.1; JOINED.  
 DR EMBL; U34922; AAC2902.1; JOINED.  
 DR EMBL; U34923; AAC2902.1; JOINED.  
 DR EMBL; U34924; AAC2902.1; JOINED.  
 DR EMBL; U34925; AAC2902.1; JOINED.  
 DR EMBL; U34926; AAC2902.1; JOINED.  
 DR EMBL; U34927; AAC2902.1; JOINED.  
 DR EMBL; U34928; AAC2902.1; JOINED.  
 DR EMBL; U34929; AAC2902.1; JOINED.  
 DR EMBL; U34930; AAC2902.1; JOINED.  
 DR EMBL; U34931; AAC2902.1; JOINED.  
 DR EMBL; U34932; AAC2902.1; JOINED.  
 DR EMBL; U34933; AAC2902.1; JOINED.  
 DR EMBL; U34934; AAC2902.1; JOINED.  
 DR EMBL; U34935; AAC2902.1; JOINED.  
 DR EMBL; U34936; AAC2902.1; JOINED.  
 DR EMBL; U34937; AAC2902.1; JOINED.  
 DR EMBL; U34938; AAC2902.1; JOINED.  
 DR EMBL; U34939; AAC2902.1; JOINED.  
 DR EMBL; U34940; AAC2902.1; JOINED.  
 DR EMBL; U34941; AAC2902.1; JOINED.  
 DR EMBL; U34942; AAC2902.1; JOINED.  
 DR EMBL; U34943; AAC2902.1; JOINED.  
 DR EMBL; U34944; AAC2902.1; JOINED.  
 DR EMBL; U34945; AAC2902.1; JOINED.  
 DR EMBL; U34946; AAC2902.1; JOINED.  
 DR EMBL; U34947; AAC2902.1; JOINED.  
 DR EMBL; U34948; AAC2902.1; JOINED.  
 DR EMBL; U34949; AAC2902.1; JOINED.  
 DR EMBL; U34950; AAC2902.1; JOINED.  
 DR EMBL; U34951; AAC2902.1; JOINED.  
 DR EMBL; U34952; AAC2902.1; JOINED.  
 DR EMBL; U34953; AAC2902.1; JOINED.  
 DR EMBL; U34954; AAC2902.1; JOINED.  
 DR EMBL; U34955; AAC2902.1; JOINED.  
 DR EMBL; U34956; AAC2902.1; JOINED.  
 DR EMBL; U34957; AAC2902.1; JOINED.  
 DR EMBL; U34958; AAC2902.1; JOINED.  
 DR EMBL; U34959; AAC2902.1; JOINED.  
 DR EMBL; U34960; AAC2902.1; JOINED.  
 DR EMBL; U34961; AAC2902.1; JOINED.  
 DR EMBL; U34962; AAC2902.1; JOINED.  
 DR EMBL; U34963; AAC2902.1; JOINED.  
 DR EMBL; U34964; AAC2902.1; JOINED.  
 DR EMBL; U34965; AAC2902.1; JOINED.  
 DR EMBL; U34966; AAC2902.1; JOINED.  
 DR EMBL; U34967; AAC2902.1; JOINED.  
 DR EMBL; U34968; AAC2902.1; JOINED.  
 DR EMBL; U34969; AAC2902.1; JOINED.  
 DR EMBL; U34970; AAC2902.1; JOINED.  
 DR EMBL; U34971; AAC2902.1; JOINED.  
 DR EMBL; U34972; AAC2902.1; JOINED.  
 DR EMBL; U34973; AAC2902.1; JOINED.  
 DR EMBL; U34974; AAC2902.1; JOINED.  
 DR EMBL; U34975; AAC2902.1; JOINED.  
 DR EMBL; U34976; AAC2902.1; JOINED.  
 DR EMBL; U34977; AAC2902.1; JOINED.  
 DR EMBL; U34978; AAC2902.1; JOINED.  
 DR EMBL; U34979; AAC2902.1; JOINED.  
 DR EMBL; U34980; AAC2902.1; JOINED.  
 DR EMBL; U34981; AAC2902.1; JOINED.  
 DR EMBL; U34982; AAC2902.1; JOINED.  
 DR EMBL; U34983; AAC2902.1; JOINED.  
 DR EMBL; U34984; AAC2902.1; JOINED.  
 DR EMBL; U34985; AAC2902.1; JOINED.  
 DR EMBL; U349



Query Match 54.7%; Score 529; DB 1; Length 1388;  
 Best Local Similarity 55.9%; Pred. No. 1.3e-43; Matches 99; Conservative 30; Mismatches 44; Indels 4; Gaps 1;

QY 7 QPVHLVALNSPLISGGMRGIRGADFOCQQARAVGLAGTFRALSSLRLQDLYSIVRADR 66  
 Db 1215 KPAHLAALNMPFSGDIR---ADFOCQQARAVGLAGTFRALSSLRLQDLYSIVRADR 66

QY 67 AAVPIVILKDELLFPSWEALFSGSEGPKPGARIFSFSDGKDGVLRHPWPKSWHGSDDP 126  
 Db 1271 YSLIPIVNLKGQVLFNNMDSIFSFGHGGOFNMHTPIYSFDGRDMDPSWPQKVWHSPPH 1330

QY 127 GRRITESYCETWRTEAATSAQASSLUGRLIGQSAASCHHAYIVLCIENSMTASK 183  
 Db 1331 GYLFLVDNCCEAWRTADTAVTGLASPLSTGKILQDKAVCANRLIVLICIENSMTARD 1387

RESULT 4  
**GLI\_CHICK**  
 ID GLI\_CHICK STANDARD: PRY; 556 AA.  
 AC P55878;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Zinc finger protein GLI (GLI) (Fragment).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
 OC Gallus.  
 NCBI\_TAXID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A. Pubmed=8948590;  
 RX MEDLINE=9705842;  
 RA Marigo V., Johnson R.L., Vortkamp A., Tabin C.J.;  
 RT "Sonic hedgehog differentially regulates expression of GLI and GLI3  
 during limb development.";  
 RL Dev. Biol. 180:273-283(1996).  
 CC -1- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING  
 NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT  
 AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL  
 NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE  
 CC -1- TRANSDUCTION OF SHH SIGNAL (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 PROTEINS.....  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC .....  
 EMBL: U60762; AAB51659.1; -.  
 DR HSPB; P08151; 2GII.  
 DR InterPro; IPR000522; Znf-C2H2.  
 PRAM; PR00096; zf-C2H2; 5.  
 SMART; SM00355; 2nfFC2H2; 5.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 5.  
 KW Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;  
 DOMAIN; Repeat.  
 FT 247 399 ZINC FINGERS.  
 FT 247 272 C2H2-TYPE.  
 FT ZN\_FING 280 307 C2H2-TYPE.  
 FT ZN\_FING 313 337 C2H2-TYPE.  
 FT ZN\_FING 343 368 C2H2-TYPE.  
 FT ZN\_FING 374 399 C2H2-TYPE.  
 FT NON\_TER 556 556 C2H2-TYPE.

SEQUENCE 556 AA; 60215 MW; 722D2AA5A1CA4D98 CRC64;

RESULT 5  
**Y4BG\_RHISN**  
 ID Y4BG\_RHISN STANDARD: PRY; 271 AA.  
 AC P55374;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-Nov-1997 (Rel. 35, Last sequence update)  
 DT 01-Nov-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical protein Y4BG precursor.  
 GN Y4BG.  
 OS Rhizobium sp. (strain NGR234).  
 OG Plasmid sym pNGR234a.  
 OC Bacterid; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 NCBI\_TAXID=394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; Pubmed=9163424;  
 RA Freiberg C.A., Fellay R., Balroch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
 RL Nature 387:394-401(1997).  
 CC -1- SIMILARITY: NONE OBVIOUS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration -  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC .....  
 EMBL; AE00068; AAB91622.1; -.  
 DR KW Hypothetical protein; Plasmid; Signal.  
 DR SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 271 HYPOTHETICAL PROTEIN Y4BG.  
 SQ SEQUENCE 271 AA; 30034 MW; 5572CF1461FC6A63 CRC64;

Query Match 7.8%; Score 75; DB 1; Length 271;  
 Best Local Similarity 21.1%; Pred. No. 3.9; Matches 28; Conservative 26; Mismatches 35; Indels 44; Gaps 7;

QY 16 NSPLSGGMMRGIRGADFOCQQARAVGLAGTFRALSSLRLQDLYSIVRA-----DRAW 69  
 Db 174 NADIAATIKSLLGADFEEAQAITGTGSGEFKS-----DDY-IGRVCTPHMCQEADL 225

QY 70 PIVNLKDELLFPSWEALFSGSEGPKPGARIFSFSDGKDGVLRHP---TWFQKSWHGSDDP 126  
 Db 226 LFLSAKDRRAYA-----Kp-----HQKKIIVRHPVKQWPEKA----- 259

QY 127 GRRITESYCETWRTEAATSAQASSLUGRLIGQSAASCHHAYIVLCIENSMTASK 139





RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurakawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 0157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 [6]  
 RN SIMILARITY TO GPPA.  
 RX MEDLINE=94025037; PubMed=8212131;  
 RA Reizer J., Reizer A., Sauer M.H. Jr., Bork B., Sander C.;  
 RT "Exopolyphosphate phosphatase and guanosine pentaphosphate:  
 Trends Biochem. Sci. 18:247-248(1993)."  
 -I FUNCTION: DEGRADATION OF INORGANIC POLYPHOSPHATES. ORTHOPHOSPHATE  
 IS RELEASED PROGRESSIVELY FROM THE ENDS OF POLYPHOSPHATE OF CIRCA  
 500 RESIDUES LONG, BUT CHAINS OF CIRCA 15 RESIDUES COMPete POORLY  
 WITH POLYPHOSPHATE AS SUBSTRATE.  
 -I CATALYTIC ACTIVITY: (polyphosphate)(N) + H(2)O =  
 [polyphosphate](N-1) + phosphate.  
 -I COPACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.  
 CC -I SUBUNIT: HOMODIMER.  
 CC -I SUBCELLULAR LOCATION: Membrane-associated.  
 CC -I SIMILARITY: BELONGS TO THE GPPA / PPX FAMILY.  
 CC  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; L06129; AAA24415.1; -.  
 DR EMBL; AE000336; AAC5555.1; -.  
 DR EMBL; D90870; BAB16390.1; -.  
 DR EMBL; D90880; BAB16392.1; -.  
 DR EMBL; AE005479; AAQ57612.1; -.  
 DR EMBL; AP002561; BAB36787.1; -.  
 DR EcoGene; EG11403; PPX.  
 DR InterPro; IPR003695; Ppx-GPPA.  
 DR Pfam; PF02241; Ppx-GPPA. 1.  
 KW Hydrolase; Magnesium; Membrane; Complete proteome.  
 FT INIT\_MET 0  
 SQ SBQUENCE 512 AA; 58004 MW; 48611AFF5D9FB9C3 CRC64;  
 Query Match 7.5%; Score 73; DB 1; Length 512;  
 Best Local Similarity 21.9%; Pred. No. 13;  
 Matches 46; Conservative 28; Mismatches 66; Indels 70; Gaps 11;  
 OY 34 PQQR-----AVLAGTGFRA-----PLSLRQLDY 58  
 DB 186 FQARMAAQKLETLTWFQRIQGWNVAMGASGTIIKAHEVLMEMGEKGDTITPPLERKLV 245  
 OY 59 -SIVRRADRAAVPVNLKDE-LLFSEWEALFG-----SEGPKPGARIFSF 103  
 DB 246 KEVLRRHNFASLSPGLSEERKTVFVGATLGCVGFDALAIRELRLSDGAIREGV-LYEM 304  
 OY 104 DCK----DVLRLPTWPKSVNGSDPGRRL---TESYCETWRTEAPS-ATGQASSLLG- 154  
 DB 305 EGRFRHDYDVRRTASSLANQYHIDSQARRVLDTQMYEOWRREQQPKLHQPLEALLRW 364  
 OY 155 -----GRLGQSAAASCHAYVILCIENS 177  
 DB 365 AAMLHEVGLNINHSGLURHSAYT---LQNS 391

AC 050654; Q50731;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 95.4 kDa protein RV2567.  
 GN RV2567 OR MT2643 OR MTCY227.34C OR MTCY9C4.01C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OC Actinomycetes; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1] NCBI\_TaxID=1773;  
 RP SEQUENCE FROM N.A.  
 RC SPRAINFH37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeyer K., Gas C., Barry C.E. III, Tekla F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R., Devlin J., Feilwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osbourn J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwynn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Brumback M.A., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Gill J., Mikl J.,  
 RA Bishai W.;  
 RT Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
 CC 6803 SLU0335.  
 CC  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; Z77250; CAB01053.1; -.  
 DR EMBL; AE007098; AAK46956.1; -.  
 DR TIGR; MT2643; -.  
 DR Tuberculist; RV2567; -.  
 KW Hypothetical protein; Complete proteome.  
 FT CONFLICT; 645 Q->R (IN REF. 2).  
 SO SEQUENCE 884 AA; 95448 MW; 95D23A4D2EDEB65 CRC64;

Query Match 7.5%; Score 72.5; DB 1; Length 884;  
 Best Local Similarity 29.5%; Pred. No. 27;  
 Matches 31; Conservative 15; Mismatches 26; Indels 33; Gaps 6;  
 OY 63 RADRAAVPVNLKDELFPSWALFSGSEGLPKGRIFSDG-----KDVLRHPTW 114  
 DB 609 RADMIAYA-----PS-TLWSLTVDPRGSLVQSVEGIALLAQVNRDQLSNDTW 656  
 OY 115 -----POKSVMGSDPGRRLTESYCETWRTEAPSMTQGASSLLG 154  
 DB 657 MYLAVNERVERHKSDP---POSLAE--ADAVLASRQETLAG 693

RESULT 11  
 CK\_HUMAN ID CRK\_HUMAN STANDARD; PRT; 304 AA.  
 AC P46108;

RESULT 10  
 YP67\_MYCTU ID CRK\_HUMAN STANDARD; PRT; 884 AA.  
 ID P46107;

DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Proto-oncogene C-crk (P38) (Adapter molecule crk).  
 GN CRK.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.;  
 RC TISSUE="Embryonic lung, and Placenta;"  
 RX MEDLINE=92334347; PubMed=1630456;  
 RA Matsuda M., Tanaka S., Nagata A., Kurata T., Shibuya M.;  
 RT "Two species of human cDNA encode proteins with distinct  
 biological activities";  
 RL Mol. Cell. Biol. 12:3482-3489 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.;  
 RC MEDLINE=93390962; PubMed=8378094;  
 RA Floretoes T., Heisterkamp N., Groffen J., Benjes S., Morris C.;  
 RT "CRK proto-oncogene maps to human chromosome band 17p13.";  
 RL Oncogene 8:2853-2855 (1993).  
 -1- FUNCTION: THE C-TERMINAL SH3 DOMAIN FUNCTION AS A NEGATIVE MODULATOR  
 ACTIVITIES. CRK-II HAS LESS TRANSFORMING ACTIVITY THAN CRK-I. BOTH  
 CRK-I AND CRK-II BIND TO MANY OF TYROSINE-PHOSPHORYLATED PROTEINS  
 THAT BIND TO GRB2.  
 -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; CRK-I AND CRK-II (SHOWN HERE);  
 ARE PRODUCED BY ALTERNATIVE SPlicing. THE CRK-I FORM LACKS THE  
 LAST 100 RESIDUES.  
 -1- DOMAIN: THE C-TERMINAL SH3 DOMAIN FUNCTION AS A NEGATIVE MODULATOR  
 FOR TRANSFORMATION AND THE N-TERMINAL SH3 DOMAIN APPEARS TO  
 FUNCTION AS A POSITIVE REGULATOR FOR TRANSFORMATION.  
 -1- PTM: PHOSPHORYLATION OF CRK-II (40 kDa) GIVES RISE TO A 42 kDa  
 FORM.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.  
 CC  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL: D10656; BAA01505.1; -;  
 DR EMBL: S65701; AAB28213.1; -;  
 DR HSPB; 064010; ICKA.  
 DR SWISS-2DPAGE; P46108; HUMAN.  
 DR MIN; 164762; -.  
 DR InterPro; IPR000980; SH2.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00017; SH2; 1.  
 DR Pfam; PRO0018; SH3; 1.  
 DR PRINTS; PRO0401; SH2DOMAIN.  
 DR SMART; SM00252; SH2; 1.  
 DR SMART; SM00326; SH3; 2.  
 DR PROSITE; PS50001; SH2; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Proto-oncogene; SH2 domain; SH3 domain; Repeat; Alternative splicing;  
 KW Phosphorylation.  
 FT DOMAIN 13 118 SH2.  
 FT DOMAIN 132 192 SH3 1.  
 FT DOMAIN 256 296 SH3 2.  
 FT VARSPLIC 205 304 MISSING (IN ISOFORM CRK-I).  
 SQ SEQUENCE 304 AA; 33872 MW; D74A83ED1FFC0EBC CRC64;

Query Match Score 72; DB 1; Length 304; Best Local Similarity 27.8%; Pred. No. 8.8; Matches 20; Conservative 12; Mismatches 24; Indels 16; Gaps 3;

RESULT 12  
 CRK\_MOUSE STANDARD PRT; 304 AA.  
 ID CRK\_MOUSE  
 AC 064110;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Proto-oncogene C-crk (P38) (Adapter molecule crk).  
 GN CRK OR CRKO.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.;  
 RC TISSUE=Liver;  
 RX MEDLINE=94239744; PubMed=8183562;  
 RA Ogawa S., Toyoshima H., Kozutsumi H., Hagiwara K., Sakai R.,  
 RA Tanaka T., Hirano N., Mano H., Yazaki Y., Hirai H.;  
 RT "The C-terminal SH3 domain of the mouse c-Crk protein negatively  
 regulates tyrosine phosphorylation of Crk associated p130 in rat 3Y1  
 cells.";  
 RL Oncogene 9:1669-1678 (1994).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 134-190.  
 RX MEDLINE=95253621; PubMed=7735837;  
 RA Wu X., Knudsen B., Feller S.M., Zheng J., Sali A., Cowburn D.,  
 RA Handa H., Kurian J.;  
 RT "Structural basis for the specific interaction of lysine-containing  
 proline-rich peptides with the N-terminal SH3 domain of c-Crk.";  
 RL Structure 3:215-226 (1995).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 133-191.  
 RX MEDLINE=9906928; PubMed=9851931;  
 RA Nguyen J.T., Turck C.W., Cohen F.E., Zuckermann R.N., Lim W.A.;  
 RT "Exploiting the basis of proline recognition by SH3 and WW domains:  
 design of N-substituted inhibitors.";  
 RL Science 282:208-2092 (1998).  
 CC -1- FUNCTION: THE CRK-I AND CRK-II FORMS DIFFER IN THEIR BIOLOGICAL  
 ACTIVITIES. CRK-II HAS LESS TRANSFORMING ACTIVITY THAN CRK-I. BOTH  
 CRK-I AND CRK-II BIND TO MANY OF TYROSINE-PHOSPHORYLATED PROTEINS  
 THAT BIND TO GRB2.  
 -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; CRK-I AND CRK-II (SHOWN HERE);  
 ARE PRODUCED BY ALTERNATIVE SPlicing. THE CRK-I FORM LACKS THE  
 LAST 100 RESIDUES.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
 CC -1- DOMAIN: THE C-TERMINAL SH3 DOMAIN FUNCTION AS A NEGATIVE MODULATOR  
 FOR TRANSFORMATION AND THE N-TERMINAL SH3 DOMAIN APPEARS TO  
 FUNCTION AS A POSITIVE REGULATOR FOR TRANSFORMATION (BY  
 SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.  
 CC  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; S72408; AAB30755.1; -.  
 DR PDB; 1CKA; 08-MAY-95.

OY 100 IFSFDGK-----DYLRRHPTWPKQSVWHGSDPNGHR-LTESYCETWRTEAPSNG 147  
 Db 140 LFDFNGNDEEDLPFKKGDPDLRKIRDPEECWWNAEDSEGKRGMPVPYVEKYR-- PASA 195  
 OY 148 QASSLIGGGLG 159  
 Db 196 SVSALLGGHQEG 207



CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR  
 DR EMBL; X97723; CAA66308.1; -.  
 DR HSSP; P08107; IHUO.  
 DR InterPro; IPR001023; HSP70.  
 DR PRINTS; PR00012; HSP70; 1.  
 DR PROSITE; PS00297; HSP70\_1; 1.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 DR  
 KW Heat shock; ATP-binding; Multigene family; Protein biosynthesis.  
 SQ SEQUENCE 613 AA; 66432 MW; 5FFBA9F8F9327F9 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 613;  
 Best Local Similarity 25.6%; Pred. No. 20; Mismatches 50; Indels 26; Gaps 6;  
 Matches 33; Conservative 20; Mismatches 50; Indels 26; Gaps 6;

Qy 44 GTFRAFLSSRLDDLYSIVRRAAVPIVNLKDELPSWEALFS-----GSGPL 94  
 Db 4 GVFQGAIGIDLGTYSCVATDSSAVETIANEQNRVPSFVFTSEERLIGDAKNGAAL 63

Qy 95 KGCAIT-----STDGKDVLRH-TWMPQSKSYWHGSDPNRRRLTE-SYCTWREAPS 144  
 Db 64 NPKKNTVFDAKRLIGRAFDDESVOKDIKSWPKKV---ESNGQPLIEVEYLDETKFSPQ 119

Qy 145 ATGQASSIL 153

Db 120 --EISSMV 125

RESULT 15

NRPL\_CHICK  
 ID NRPL\_CHICK STANDARD; PRM; 312 AA.

AC P55806;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE NAD(P)(+)-arginine ADP-ribosyltransferase 1 precursor (EC 2.4.2.31)  
 DE (Mono-ADP-ribosyl)transferase 1 (AT1).  
 OC Gallus gallus (chicken)  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 RN NCBI\_TaxID=9031;  
 [1]  
 RP  
 RP SEQUENCE FROM N. A., AND PARTIAL SEQUENCE.  
 RC STRAIN=WHITE LEGHORN; TISSUE=Bone marrow;  
 RX MEDLINE=95030487; PubMed=961658;  
 RA Tsuchiya M., Hara N., Osago H., Shimoyama M.;  
 RT "Cloning and expression of cDNA for arginine-specific ADP-  
 ribosyltransferase from chicken bone marrow cells.";  
 RL J. Biol. Chem. 269:27451-27457(1994).  
 CC -!- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N2-  
 CC (ADP-D-ribosyl)-L-arginine.  
 CC -!- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED  
 CC FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.  
 CC -!- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE  
 CC FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR  
 DR EMBL; D31864; BAA06664.1; -.  
 DR InterPro; IPR000708; ART.  
 DR Pfam; PF01129; ART; 1.  
 DR PRINTS; PR00070; RIBTRANSFRASE.  
 DR PROSITE; PS01291; ART; 1.  
 DR PROTEP; 21 31  
 FT CHAIN 32 256  
 FT SIGNAL 1 20  
 FT TRANSFERASE; Glycosyltransferase; NAD; Signal; Zymogen..  
 FT PROPEP 267 312  
 FT ACT\_SITE 224 224  
 SQ SEQUENCE 312 AA; 35318 MW; B82980439BC904FC CRC64;

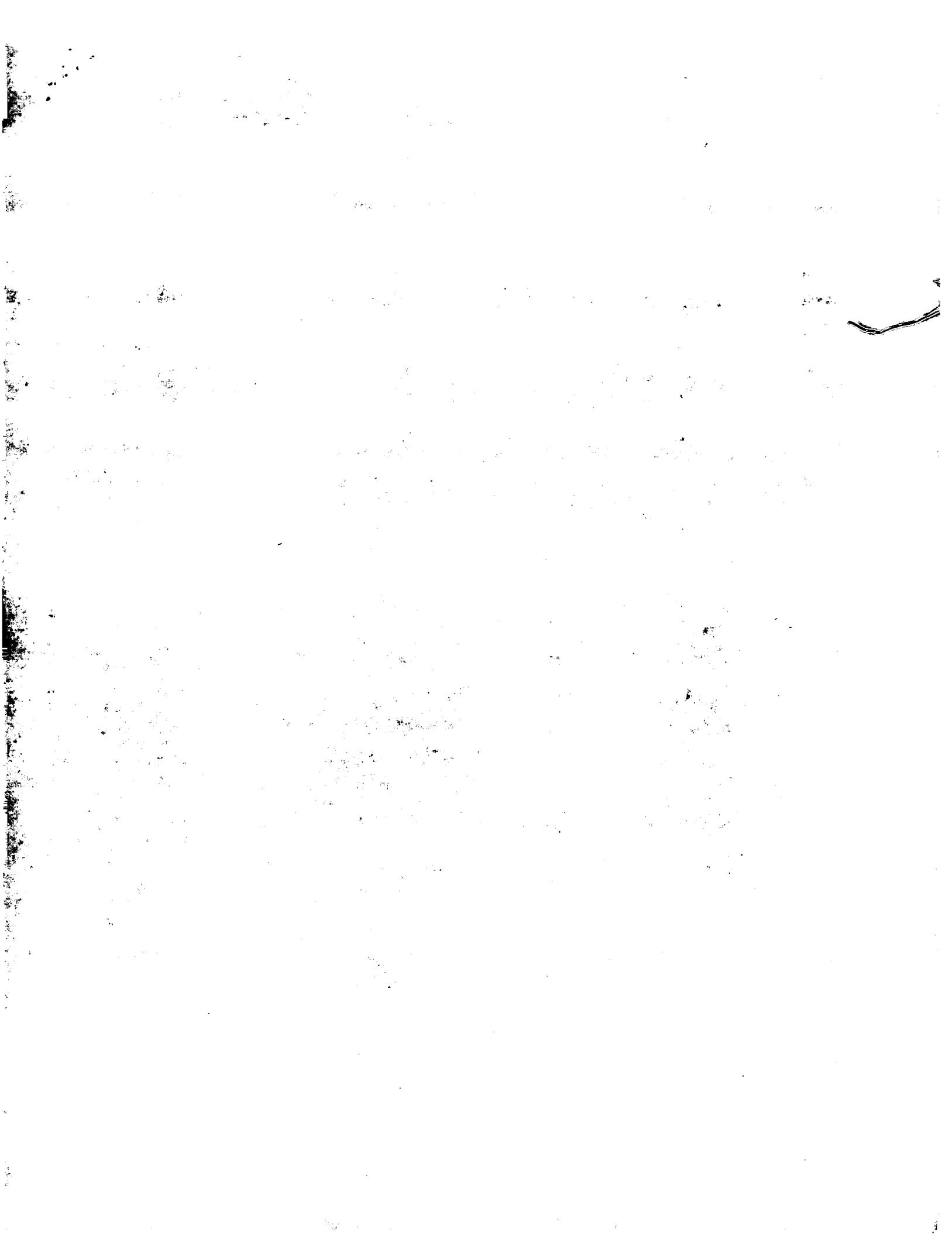
Query Match 7.3%; Score 70.5; DB 1; Length 312;  
 Best Local Similarity 24.8%; Pred. No. 13; Mismatches 45; Indels 29; Gaps 5;  
 Matches 30; Conservative 17; Mismatches 45; Indels 29; Gaps 5;

Qy 2 SHRDQPVHLVALNSPLGGMRGIRGADFQCFQOARAV-----GLACTPRAFLS 52  
 Db 128 SHQDYIHSYHKTLHFLLTQLFALFALARASOPRCYVYRGIRFMQTKSVRFQFTST 187

Qy 53 RL-----QDLYSIVRRAAVPIVNL----KELLPSPWEAL---FSGSGPL 94  
 Db 188 SLRKETAVNFGQDTLFVYVTC--YGVPIKQFSFPSEDEVLLIPPEEVFEVINFSDRGV 245

Qy 95 K 95  
 Db 246 K 246

Search completed: August 12, 2002, 10:14:43  
 Job time: 130 sec







A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous  
A;Reference number: A58371; MUID:94240112  
A;Accession: A55371  
A;Molecule type: mRNA  
A;Residues: 1-928 <RER2>  
A;Cross-references: GB:Li6898; NID:9404754; PIDN:AAA37434.1; PID:9553894  
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
R;Oh, S.P.; Kamagata, Y.; Murabak, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
R;Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-xaa  
A;Reference number: A58370; MUID:94240111  
A;Accession: S6595  
A;Molecule type: mRNA  
A;Residues: 28-1315 <OHS>  
A;Cross-references: EMBL:122545  
R;Oh, S.P.; Kamagata, Y.; Murabak, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
A;Title: Reference number: A58370; MUID:94240111  
A;Accession: S6595  
A;Molecule type: mRNA  
A;Cross-references: EMBL:122545; NID:9348968; PIDN:AAA19787.1; PID:9511298  
R;Oh, S.P.; Kamagata, Y.; Murabak, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-xaa  
A;Reference number: A58370; MUID:94240111  
A;Accession: S6595  
A;Molecule type: mRNA  
A;Residues: 28-1315 <OHS>  
A;Cross-references: EMBL:122545  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit are  
lated and subsequently O-glycosylated.  
C;Comment: The different splice forms of collagen alpha 1(XVIII) chain by the action of enzymes may be useful in treating solid tumors.  
C;Superfamily: unassigned collagens  
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglycans  
F;1-25/Domain: signal sequence #status predicted <SIG>  
F;2-235/Region: thrombospondin amino-terminal similarity  
F;2-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <MAP>  
A;Map position: 10:41.0  
F;3-27-353/Domain: collagenous #status predicted <COI>  
F;3-64-437/Domain: collagenous #status predicted <CO2>  
F;6-58-/Domain: collagenous #status predicted <CO3>  
F;6-67-689-/Domain: collagenous #status predicted <CO4>  
F;7-94-745/Domain: collagenous #status predicted <CO5>  
F;7-59-831-/Domain: collagenous #status predicted <CO6>  
F;8-42-874/Domain: collagenous #status predicted <CO7>  
F;8-87-910/Domain: collagenous #status predicted <CO8>  
F;8-92-894/Region: cell attachment (R-G-D) motif  
F;9-18-969/Domain: collagenous #status predicted <CO9>  
F;9-93-1000/Domain: collagenous #status predicted <CO10>  
F;11-32-1315/Region: endostatin #status predicted <EST>  
F;11-139-1315/Region: multiplexin collagen carboxyl-terminal similarity  
F;12-6-488/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;17-2-228/Disulfide bonds: #status predicted  
F;24-0-245-1257/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F;45-1-454-594/Binding site: chondroitin sulfate (ser) (covalent) #status predicted

Query Match 86.6%; Score 837; DB 2; Length 1315;  
Best Local Similarity 85.1%; Pred. No. 6, 4e-75;  
Matches 154; Conservative 16; Mismatches 11; Indels 0; Gaps 0;  
Qy 1 HSHRDPQLVHLVALNPLSPLSGMGRGIRGAQDFOCQQARAVGLAGTFRALFLSSRLQDIYSI 60  
Db 1132 HTHQDFQPQVVLHLVALNPLSPLSGMGRGIRGAQDFOCQQARAVGLAGTFRALFLSSRLQDIYSI 1191  
Qy 61 VRRADRAAVPVTNLKDLFLPSWAEALFGSEGPPLKGPARISFFDGKVLRHPTWPQSKVW 120  
Db 1192 VRRADRGSPVTPNLKDLFLPSWDSLPSQGQVQCARIFSEDFGRVLRHPTWPQSKVW 1251  
Qy 121 HGSDPNSGRRLRTESYCYEWRTEAPASQGASSLGGRLGOSAASCHAYIVLCTENSFMT 180  
Db 1252 HGSDPNSGRRLMESYCETWRTEATGQASSLISGRLEQKASCHNSYIVLCTENSFMT 1311  
Qy 181 A 181  
Db 1312 S 1312

RESULT 4  
A53317  
collagen alpha 1(XV) chain precursor - human  
C;Species: Homo sapiens (man)  
C;Alternate names: procollagen alpha 1(XV) chain  
C;Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 31-Mar-2000  
C;Accession: A53317; A53146; S28778  
R;Kivilikko, S.; Heinaemaki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, J. Biol. Chem. 269, 4773-4779, 1994  
A;Title: Primary structure of the alpha1 chain of human type XV collagen and exon-int  
A;Reference number: A53317; MUID:94148920  
A;Accession: A53317  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1388 <KIT>  
A;Cross-references: GB:125280  
A;Note: nucleotide sequence and conceptual translation not complete  
R;Muragaki, Y.; Abe, N.; Niinomiya, Y.; Olsen, B.R.; Ooshima, A.  
J. Biol. Chem. 269, 4042-4046, 1994  
A;Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triplet repeat domain.  
A;Reference number: A53146; MUID:94140817  
A;Accession: A53146  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-9 / 'S' / 11-48, 'V' / 50-94, 'A' / 95-149, 'A' / 151-203, 'V' / 205-408, 'A' / 410-569 <MAP>  
A;Cross-references: GB:D21220; NID:9415605; PIDN:BA04762.1; PID:di00524; PID:946070  
R;Myers, J.C.; Kivilikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992  
A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), characterized by a long proline-rich domain.  
A;Reference number: S28778; MUID:93066196  
A;Accession: S28778  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 544-640, 'P' / 642-811, 'P' / 813-1252 <MYE>  
C;Genetics:  
A;Gene: GDB:COL15A1  
A;Cross-references: GDB:132578; OMIM:120325  
A;Map position: 9q21-9q22  
C;Superfamily: unassigned collagens  
F;1-22/Domain: signal sequence #status predicted <SIG>  
F;23-188/Product: collagen alpha 1(XV) chain #status predicted <MAP>  
F;1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 54.7%; Score 529; DB 2; Length 1388;  
Best Local Similarity 55.9%; Pred. No. 3, 4e-44;  
Matches 99; Conservative 30; Mismatches 44; Indels 4; Gaps 1;  
Qy 7 QPVYLHLVALNPLSPLSGMGRGIRGAQDFOCQQARAVGLAGTFRALFLSSRLQDIYSI 66  
Db 1215 KPALHLAALNMPFSGDIR--ADFOCFQKQARAAGLSTYRAFLSSRLQDIYSI 1270  
Qy 67 AAQPIVNLKDELFLPSWALFSEGPPLKGPARISFFDGKVLRHPTWPQSKVW 126  
Db 1271 YSLPVNLKQVQFLNNWSSIFSGGGQFNMHPIYFSDFGRDMDPWSQVWVHGSSP 1330  
Qy 127 GRILUTESCLCTWRTEAPASQGASSLGGRLGOSAASCHAYIVLCTENSFMT 183  
Db 1331 GVRLVDNYCEAWRHPADTAVTGLASPLSTGKILDQKAVSCANRLIVLCIENSFTDAR 1387

RESULT 5  
T22002  
hypothetical protein F39H11.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T22002  
R;White, S.  
submitted to the EMBL Data Library, October 1999  
A;Reference number: Z10500  
A;Accession: T22002

A; Status: preliminary; translated from GB/EMBL/DBJ  
 A; Molecule type: DNA  
 A; Residues: 1-650 <WIL>  
 A; Cross-references: EMBL:281079; PIDN:CA03084.1; GSPDB:GN00019; CESP:F39H11.4  
 A; Experimental source: clone F39H11  
 C; Genetics:  
 A; Gene: CESP:F39H11.4  
 A; Map position: 1  
 A; Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 38.6%; Score 373; DB 2; Length 650;  
 Best Local Similarity 44.1%; Pred. No. 4. 7e-29;  
 Matches 75; Conservative 26; Mismatches 63; Indels 6; Gaps 4;

Oy 9 VHLHALNPLSPLGGMGIRGADFGCPOQARAVLAGTFRALSRQDLSIVRRADRAA 68  
 Db 464 VHMFLALSOPPSGNLHGLRGADLOYREARAGYTTFRAMLSSNVQDLRIVHSVD-FD 522

Oy 69 VPIVNLKDELLFSPMSMEALFGSGEPLPGKARIFSFQDGKDVLRLHPTWPQKSWMHGSDP-NGR 128  
 Db 523 TTIVNVNAGHHLFSPNSRFVNGAQ--MNPFLAKLFSDFRDVYLNDSRWPDKRVWGSKDGGI 580

Oy 129 RLTESYCETWRTEASATGQASSLIGRLGQSAMS -CHHAYVLCIEN 176  
 Db 581 R-AEQCDGWRRADSLTLSLGHISNTSIFQSGSEKCECNKLWVLCVEN 629

RESULT 6

T05202 pectinesterase homolog F4110\_150 - *Arabidopsis thaliana*  
 C; Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C; Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999  
 C; Accession: T05202  
 R; Bevan, M.; Peters, S.A.; van Staveren, M.; Dirks, W.; Stiekema, W.; Hohisel, J.; Meijer submitted to the Protein Sequence Database, February 1999  
 A; Reference number: 215402  
 A; Accession: T05202  
 A; Molecule type: DNA  
 A; Residues: 1-777 <REV>  
 A; Cross-references: EMBL:AL039525  
 A; Experimental source: cultivar Columbia; BAC clone F4110  
 C; Genetics:  
 A; Map position: 4  
 A; Introns: 7/3; 47/1; 90/1; 246/1; 354/1  
 A; Note: F4110\_150  
 C; Superfamily: pectinesterase

Query Match 8.2%; Score 79; DB 2; Length 477;  
 Best Local Similarity 21.9%; Pred. No. 6.5; Matches 32; Conservative 25; Mismatches 53; Indels 36; Gaps 6;

Oy 15 LNSPLSPGGMGIKGADFOCFOQARAVLAGTFRALSRQDLSIVRRADRAAVPIVN 74  
 Db 63 LSAALSNOATCMEGFD-----GTSLVKSILVAGSLDQLYSMLRE---LLPLVQ- 107

Oy 75 KDELLEPWSWEALFSESEGPL----KPGARIFSFQDGKDVLRLHPTWPQKSWMHGSDP-NGR 128  
 Db 108 -----PEQPKAVSKPGPKTAKGPKAPPGKHLRDRDEDSLQFPWVR-----PDDR 153

Oy 129 RLTESYCETWRTE-APSATQOASLL 153  
 Db 154 KLLESNGRTDVSDLGFTKIM 179

RESULT 7

F98231 succinoglycan biosynthesis protein exom [imported] - *Agrobacterium tumefaciens* (strain C)  
 C; Species: *Agrobacterium tumefaciens*  
 C; Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
 Q; Accession: F98231  
 R; Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, Science 294, 2323-2328, 2001  
 A; Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium*  
 A; Reference number: A97359; PMID:11743194  
 A; Accession: F98231  
 A; Status: preliminary  
 A; Residues: 1-314 <KUR>  
 A; Cross-references: GB:AE007870; PIDN:AAK89376.1; PID:q15159227; GSPDB:GN00170  
 A; Gene: AGR\_L1606  
 A; Map position: linear chromosome C; Genetics:  
 C; Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

Query Match 8.0%; Score 77; DB 2; Length 314;  
 Best Local Similarity 27.9%; Pred. No. 6.1; Matches 29; Conservative 14; Mismatches 41; Indels 20; Gaps 5;

Oy 76 DELFSPSW-----EALFSGEPLPGKARIFSFQDGKDVLRLHPTWPQKSWMHGSDP--N 126  
 Db 105 DETAPPHWLALLEETETQETVLPQPVITVWRDN-----PGMKRGDFHSTPVWVN 158

Oy 127 GRLTESYCET-WRTEAPSATGQASSLIGRLGQSASAASHHAY 169  
 Db 159 GEITIGTYCTNLLRMEAPVKRRFAL---ALGQSGGEDTHFF 198

RESULT 8

AG3054 succinoglycan biosynthesis protein exom [imported] - *Agrobacterium tumefaciens* (strain C)  
 C; Species: *Agrobacterium tumefaciens*  
 C; Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
 C; Accession: AG3054  
 R; Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wooldridge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.; Ster, E.W.  
 A; Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A; Reference number: AB2577; PMID:11743193  
 A; Accession: AG3054  
 A; Molecule type: DNA  
 A; Residues: 1-314 <KUR>  
 A; Cross-references: GB:AE008699; PIDN:AAA44853.1; PID:q17742499; GSPDB:GN00187  
 A; Experimental source: strain C58 (Dupont)  
 C; Genetics:  
 A; Map position: preliminary  
 A; Gene: exom  
 A; Residues: 1-314 <KUR>  
 C; Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

Query Match 8.9%; Score 77; DB 2; Length 314;  
 Best Local Similarity 27.9%; Pred. No. 6.1; Matches 29; Conservative 14; Mismatches 41; Indels 20; Gaps 5;

Oy 76 DELFSPSW-----EALFSGEPLPGKARIFSFQDGKDVLRLHPTWPQKSWMHGSDP--N 126  
 Db 105 DETAPPHWLALLEETETQETVLPQPVITVWRDN-----PGMKRGDFHSTPVWVN 158

Oy 127 GRLTESYCET-WRTEAPSATGQASSLIGRLGQSASAASHHAY 169  
 Db 159 GEITIGTYCTNLLRMEAPVKRRFAL---ALGQSGGEDTHFF 198

RESULT 9

B45022 CRK I - human  
 C; Species: Homo sapiens (man)  
 C; Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 12-Feb-1999  
 C; Accession: B45022

R;Matsuda, M.; Tanaka, S.; Nagata, S.; Kojima, A.; Kurata, T.; Shibuya, M.  
Mol. Cell. Biol. 12, 3482-3489, 1992  
A;Title: Two species of human CRK cDNA encode proteins with distinct biological activities  
A;Reference number: A45022; MUID:92334347

A;Accession: B45022  
A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-204 <MAY>

A;Experimental source: embryonic lung cells

A;Note: sequence extracted from NCBI backbone (NCBIN:108771, NCBIP:108772)

C;Superfamily: ckr transforming protein; SH2 homology; SH3 homology

F:139-187/Domain: SH2 homology <SH2>

F:139-187/Domain: SH3 homology <SH3>

Query Match 7.9%; Score 76; DB 2; Length 204;  
Best Local Similarity 29.0%; Pred. No. 4.6; Mismatches 20; Indels 16; Gaps 3;  
Matches 21; Misnatches 21; Indels 16; Gaps 3;

QY 100 IFSPFDGK-----DVLRHPTWPOKSVMWHGSDPGR--LITESVCETWRPEAPSATG 147  
Db 140 LDFNGNDEEDLFPKKGDLIRRDKPEQWNAEDSEGKRGKMPVPPVEKIR---PASA 195

QY 148 QASLLIGGR 156  
Db 196 SVSVALIGGR 204

RESULT 10  
JW0047  
class I cytokinase receptor precursor - human  
N;Alternate names: WSK-1  
C;Species: Homo sapiens (man)  
C;Date: 17-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 21-Jul-2000  
C;Accession: JW0047  
R;Sprecher, C.A.; Grant, F.J.; Baumgartner, J.W.; Presnell, S.R.; Schrader, S.K.; Yamaguchi, R.; Biochem. Biophys. Res. Commun. 246, 82-90, 1998  
A;Title: Cloning and characterization of a novel class I cytokine receptor.  
A;Reference number: JW0047; MUID:98262921  
A;Accession: JW0047  
A;Molecule type: mRNA  
A;Residues: 1-636 <SPR>

A;Cross-references: GB:AF053004; NID:q3153240; PIDN: AAC39755.1; PID:g3153241

C;Genetics:

A;Map position: 19p13.11

C;Keywords: glycoprotein

F:1-32/Domain: signal sequence #status predicted <SIG>

F:515-540/Domain: transmembrane #status predicted <TM>

F:554-561/Domain: cytoplasmic #status predicted <CRP>

F:51,76,302,311,374,382,467/Binding site: carbohydrate (Asn) (covalent) #status predicted <Covalent>  
Query Match 7.9%; Score 76; DB 2; Length 636;  
Best Local Similarity 24.2%; Pred. No. 18; Matches 39; Conservative 12; Mismatches 46; Indels 64; Gaps 7;

QY 23 MRGIRGADEF-----QCFQFQARAVGLAGTFRALSSLRQLD----- 57  
Db 1 MRGGRRGAPFWLPLPKLALLPLPLWVLFQRTFGQSACPLQCYGVGPGLDNCWSPEPLGD 60

QY 58 -----YSTIVRADRAAVPI---VNLKDBLLF-----PSWEALFLS 88  
Db 61 GAPSELHLQSQRYSRNSKTQTVWAAGRSWVAIFREQLTMSDKLVLWGTAKGQPLWPPVFV 120

QY 89 GSSEGPLKFGA\_RI--FSFDGKDVLR-----HPTWQPKSV 119  
Db 121 NLLETQMPNPAPRLGPDVDFSFEDDPLEATVHWRPPTWPSHKV 161

RESULT 11  
T04377  
probable alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41) - barley

N;Alternate names: pullulanase

C;Species: Hordeum vulgare (barley)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999

C;Accession: T04377  
R;Iok, F.; Kristensen, M.; Planchot, V.; Leah, R.; Svendsen, I.; Svenson, B.

submitted to the EMBL Data Library, December 1997

A;Description: Isolation and characterization of starch debranching enzyme, limit dextrinase

A;Reference number: Z15320

A;Experimental source: EMBL:AF022725; NID:92502057; PIDN:AAD04189.1; PID:92677837

A;Molecule type: DNA

A;Gene: HvLD99

A;Introns: 21/3; 70/1; 87/3; 124/1; 169/3; 235/2; 285/2; 305/3; 346/1; 376/2; 407/3;

C;Keywords: glycosidase; hydrolase

Query Match 7.8%; Score 75.5; DB 2; Length 904;  
Best Local Similarity 22.5%; Pred. No. 31; Mismatches 45; Conservative 19; Misnatches 67; Indels 69; Gaps 8;

QY 17 SPLSGGMRCIRGADFOCFCQFQARAVGLAGT-----FRAFLSSRLQDLYSIVR--- 62  
Db 62 SPZNG---GIOQADSKVELQPEASGLPENVTQKFPLISSYRARKVPSSWVDSLVKOLV 118

QY 63 ---RADRAVPTVNL-----KDELIFPSWE-----ALFSG 89  
Db 119 VASFGADGKHVDYTGLOEFGVLDMFAYTGPLGAVFSEDSVSHLWAPTAQGVSVCFD 178

QY 90 SEGP-----LKPGARISSEFDGK-----DVLRHPTWPOKSVMWHGSUPNGRLT 131  
Db 179 PAGPALETWVOLKESNGVWNSVTVGPREWENRYLYEVDVY-HPTKAQVILKCLAGDPYARSLS 237

QY 132 ESCETWREAPSATGQASS 151  
Db 238 ANGARTWLVDDINNETLPKAS 257

RESULT 12  
B89781  
conserved hypothetical protein SA0184 [imported] - *Staphylococcus aureus* (strain N315  
C;Species: *Staphylococcus aureus*  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: B89781  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, T.; Cui, L.; Oma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 355, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: B89781  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-351 <KUR>

A;Cross-references: GB:BA000018; PID:913700106; PIDN:BAB41405.1; GSPPDB:GN00149

A;Experimental source: strain N315

C;Genetics:  
A;Gene: SA0184

Query Match 7.8%; Score 75; DB 2; Length 351;  
Best Local Similarity 23.3%; Pred. No. 11; Mismatches 42; Conservative 17; Misnatches 67; Indels 54; Gaps 8;

QY 23 MRGIRGADEFQFQARAVGLAGTFRALSSLRQLD-----YSTIVRADRAAVPIV 72  
Db 97 IEAMAQKICCCNALS----RELIISLHQLNIPFTLSCFHNVYPRPDTGLSYDLY 151

QY 73 NLKDBLLFESWEALFSGSEGPLKFGAIFSFCDKDVLRHPTWPKSTWHSGPNGRRLTE 132

Db	152	NKKNELLY--	-QNPRAIYFIVGSLRGPPL-----HKGKPT-----	186
Qy	133	SYCETWNTTEAP-----SATQASSLIGRLGOSAS---CHRAYIVCIENSFT	180	
Db	187	--IEATRHSHPVAAKLLQEQTGVSEVLVNGDSLIEMROAKQLIDFCRKHRHTLCIEEVFTD	244	
RESULT	13	gastric mucin MUC5AC - human (fragment)		
C;Species:	Homo sapiens (man)			
C;Date:	27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Apr-2000			
C;Accession:	S56015; S53361			
R;Klopm, L.W.J.; van Rens, L.; Strous, G.J.				
Blochman, J. 308, 831-838, 1995				
A;Title: Cloning and analysis of human gastric mucin cDNA reveals two types of conserved				
A;Reference number: S56015; MUID:97104281				
A;Accession: S56015				
A;Status: preliminary				
A;Molecule type: mRNA				
A;Residues: 1-850 <KLO>				
A;Cross-references: EMBL:X81649; NID:9547516; PIDN:CAA57309.1; PID:9547517				
B;Guyonnet-Duperat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buisine, M.P.; Gallegue-Blochman, J. 305, 211-219, 1995				
A;Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich domain				
A;Reference number: S53361; MUID:95126907				
A;Accession: S53361				
A;Status: preliminary; nucleic acid sequence not shown				
A;Molecule type: mRNA				
A;Residues: 648-678, L' 680-733, L' 735-760 <GY>				
A;Cross-references: EMBL:334280; NID:9563380; PIDN:CAA84034.1; PID:9563381				
A;Experimental source: clone JU132				
A;Note: this publication is not cited in GenBank entry HSMUCINS, release 113.0				
Query Match 7.7%; Score 74.5; DB 2; Length 850; Best Local Similarity 24.7%; Pred. No. 37; Matches 39; Conservative 23; Mismatches 71; Indels 25; Gaps 6; Qy 25 GIRGAPOFGQDQARAVAGLAGTERAFLSSRLDYLSTRRADR-AAVPVNLDELLIPSW 83 Db 592 GINGGDQDTQNLNRDEGY-TF----CESPRSVQCRAESFPNTPLADLGQDVICSH 642 Qy 84 EALFSSEGPKLPK-----GARIFSFQDGKDVLRLRHPTWQKSVM-----HGSPPNGRRLTE 132 Db 643 EGЛИCLINKNQLPICNYEIRIOCCETVNVCRDITRPPKTVATRTPRPHPTQAQTQTTFT 702 Qy 133 SYCETWNTTEAPSATGQ---ASSSLGRLLGOSAASCH 166 Db 703 THMPSASTEQPRTSRRGGPPTATSYGTQTHITFVTRNCH 740				
RESULT	14	hypothetical protein ECS3386 [imported] - Escherichia coli (strain 0157:H7, substrain B91052)		
C;Species:	Escherichia coli			
C;Date:	18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001			
C;Accession:	B91052			
R;Hayashi, T.; Makino, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.				
DNA Res. 8, 11-22, 2001				
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and 9				
A;Reference number: A99629; MUID:21156231; PMID:1258796				
A;Accession: B91052				
A;Status: preliminary				
A;Molecule type: DNA				
A;Residues: 1-1653 <GY>				
A;Cross-references: GB:BA000007; PIDN:BAB36809.1; PID:913362856; GSPDB:GN00154				
A;Experimental source: strain O157:H7, substrain RIMD 0509952				
C;Genetics:				
A;Gene: ECS3386				
Query Match 7.6%; Score 73.5; DB 2; Length 1653; Best Local Similarity 24.4%; Pred. No. 1e+02; Matches 32; Conservative 21; Mismatches 47; Indels 31; Gaps 5; Qy 41 GLAGTFRAFLSSRLQDLY-----STVRADDAAVPTVNLKBLLELFPSSWEAFLSGSE 91 Db 380 GAPGYSKQFMFGPRDRYRGETVILNGLIRDADOKALPNQPIKUDVIRPDQWLRSVS 439 Qy 92 GPLKPGARIFSFQDGKDVLRLRHPTWQKSVM-----GSDPNGRRLRTSYCERWRTE-- 141 Db 440 QP-----ENGLRFHTWPQPLDSNATGMWHIRANQGDNDQYRMWDPFHVEDPERM 487 Qy 142 APSATQOASSL 152 Db 488 ALNLTGKEKTPPL 498				
Search completed: August 12, 2002, 10:13:51 Job time: 128 sec				

• Query Match Score 74; DB 2; Length 1063;  
Best Local Similarity 23.3%; Pred. No. 54;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:11:08 ; Search time 13.23 Seconds  
(without alignments)  
337.860 Million cell updates/sec

Title: US-10-080-797-1  
Perfect score: 967  
Sequence: 1 HSHRDFQPVHLVALNSPLS..... SCHRAYIVLCLIENSFMTASK 183

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24125594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%, Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/iaa/PCIUS\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/iaa/Backfileseq1.pep: \*  
6: /cgn2\_6/ptodata/2/iaa/Backfileseq1.pep: \*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Length	DB ID	Description
1	967	100.0	183	4 US-09-206 059-2
2	962	99.5	182	4 US-09-561-500-14
3	962	99.5	182	4 US-09-561-108-14
4	962	99.5	182	4 US-09-315-689-3
5	962	96.9	178	4 US-09-315-689-5
6	840	86.9	191	4 US-09-561-500-13
7	840	86.9	191	4 US-09-561-108-13
8	837	86.6	195	4 US-09-159-784-2
9	796	82.3	185	3 US-08-985-526-36
10	53.4	191	1	US-08-159-784-3
11	160	16.5	35	3 US-09-046-985-2
12	160	16.5	35	4 US-09-474-743-2
13	101	10.4	22	4 US-09-46-985-2
14	101	22	4	US-03-474-743-7
15	96	9.9	20	2 US-08-740-168A-1
16	9.9	20	4	US-09-34-429-1
17	96	9.9	20	4 US-03-315-689-1
18	94	9.7	16	4 US-03-385-442-32
19	76	7.9	578	2 US-03-73-594-3
20	76	7.9	578	3 US-03-275-925-3
21	76	7.9	578	3 US-03-275-925-3
22	7.9	636	1	US-08-563-740-5
23	7.9	636	2	US-03-073-594-5
24	7.9	636	3	US-03-275-925-5
25	7.4	256	1	US-07-906-349A-8
26	7.4	256	1	US-08-167-035-4
27	7.7	256	1	US-08-208-887A-4

**ALIGNMENTS**

RESULT 1

Query Match 100.0%; Score 967; DB 4; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2e-110; 0; Mismatches 0; Indels 0; Gaps 0;

Matches 183; Conservative 0; Sequence 1, Appli  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 4, Appli  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 7, Appli  
Sequence 8, Appli  
Sequence 9, Appli  
Sequence 10, Appli  
Sequence 11, Appli  
Sequence 12, Appli  
Sequence 13, Appli  
Sequence 14, Appli  
Sequence 15, Appli  
Sequence 16, Appli  
Sequence 17, Appli  
Sequence 18, Appli  
Sequence 19, Appli  
Sequence 20, Appli  
Sequence 21, Appli  
Sequence 22, Appli  
Sequence 23, Appli  
Sequence 24, Appli  
Sequence 25, Appli  
Sequence 26, Appli  
Sequence 27, Appli  
Sequence 28, Appli  
Sequence 29, Appli  
Sequence 30, Appli  
Sequence 31, Appli  
Sequence 32, Appli  
Sequence 33, Appli  
Sequence 34, Appli  
Sequence 35, Appli  
Sequence 36, Appli  
Sequence 37, Appli  
Sequence 38, Appli  
Sequence 39, Appli  
Sequence 40, Appli  
Sequence 41, Appli  
Sequence 42, Appli  
Sequence 43, Appli  
Sequence 44, Appli  
Sequence 45, Appli  
Sequence 46, Appli  
Sequence 47, Appli  
Sequence 48, Appli  
Sequence 49, Appli  
Sequence 50, Appli  
Sequence 51, Appli  
Sequence 52, Appli  
Sequence 53, Appli  
Sequence 54, Appli  
Sequence 55, Appli  
Sequence 56, Appli  
Sequence 57, Appli  
Sequence 58, Appli  
Sequence 59, Appli  
Sequence 60, Appli  
Sequence 61, VRRADRAAPVNUKDELFLPSWEALFSGSEGPPIKPGARIKFSEGDKDVRHPTWPKOSW 120  
Db 1 HSHRDFQPVHLVALNSPLSGGMGRIGADFOCQOARAVLAGTFRAPLSSRLQDIXSI 60  
Qy 1 HSHRDFQPVHLVALNSPLSGGMGRIGADFOCQOARAVLAGTFRAPLSSRLQDIXSI 60  
Db 1 HSHRDFQPVHLVALNSPLSGGMGRIGADFOCQOARAVLAGTFRAPLSSRLQDIXSI 60  
Qy 61 VRRADRAAPVNUKDELFLPSWEALFSGSEGPPIKPGARIKFSEGDKDVRHPTWPKOSW 120  
Db 61 VRRADRAAPVNUKDELFLPSWEALFSGSEGPPIKPGARIKFSEGDKDVRHPTWPKOSW 120  
Qy 121 HGSDPNGRLLTESCTEWAPSATGQASSLAGRLQGOSAASCHAYIVLCLIENSFMT 180  
Db 121 HGSDPNGRLLTESCTEWAPSATGQASSLAGRLQGOSAASCHAYIVLCLIENSFMT 180  
Qy 181 ASK 183  
Db 181 ASK 183

RESULT 2

Sequence 1, Appli  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 4, Appli  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 7, Appli  
Sequence 8, Appli  
Sequence 9, Appli  
Sequence 10, Appli  
Sequence 11, Appli  
Sequence 12, Appli  
Sequence 13, Appli  
Sequence 14, Appli  
Sequence 15, Appli  
Sequence 16, Appli  
Sequence 17, Appli  
Sequence 18, Appli  
Sequence 19, Appli  
Sequence 20, Appli  
Sequence 21, Appli  
Sequence 22, Appli  
Sequence 23, Appli  
Sequence 24, Appli  
Sequence 25, Appli  
Sequence 26, Appli  
Sequence 27, Appli  
Sequence 28, Appli  
Sequence 29, Appli  
Sequence 30, Appli  
Sequence 31, Appli  
Sequence 32, Appli  
Sequence 33, Appli  
Sequence 34, Appli  
Sequence 35, Appli  
Sequence 36, Appli  
Sequence 37, Appli  
Sequence 38, Appli  
Sequence 39, Appli  
Sequence 40, Appli  
Sequence 41, Appli  
Sequence 42, Appli  
Sequence 43, Appli  
Sequence 44, Appli  
Sequence 45, Appli  
Sequence 46, Appli  
Sequence 47, Appli  
Sequence 48, Appli  
Sequence 49, Appli  
Sequence 50, Appli  
Sequence 51, Appli  
Sequence 52, Appli  
Sequence 53, Appli  
Sequence 54, Appli  
Sequence 55, Appli  
Sequence 56, Appli  
Sequence 57, Appli  
Sequence 58, Appli  
Sequence 59, Appli  
Sequence 60, Appli  
Sequence 61, VRRADRAAPVNUKDELFLPSWEALFSGSEGPPIKPGARIKFSEGDKDVRHPTWPKOSW 120  
; Sequence 14, Application US/09561500  
; Patent No. 6342219  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekke  
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

FILE REFERENCE: 4001-002500  
 CURRENT APPLICATION NUMBER: US/09/561,500  
 CURRENT FILING DATE: 2000-04-28  
 PRIOR APPLICATION NUMBER: 60/131,432  
 PRIOR FILING DATE: 1999-04-28  
 NUMBER OF SEQ ID NOS: 44  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 14  
 LENGTH: 182  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
 OTHER INFORMATION: PEPTIDE  
 US-09-561-500-14

Query Match 99.5%; Score 962; DB 4; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-110; Mismatches 0; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; Patents 0; Sequence 3, Application US/09315689  
 ; Patent No. 6346510  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Folkman, Judah  
 ; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions  
 ; FILE REFERENCE: 0513-0229  
 ; CURRENT APPLICATION NUMBER: US/09/315,689  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 182  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-315-689-3

RESULT 4  
 Query Match 99.5%; Score 962; DB 4; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-110; Mismatches 0; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; Patents 0; Sequence 3, Application US/09315689  
 ; Patent No. 6346510  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Folkman, Judah  
 ; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions  
 ; FILE REFERENCE: 0513-0229  
 ; CURRENT APPLICATION NUMBER: US/09/315,689  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 182  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-315-689-3

RESULT 3  
 Query Match 99.5%; Score 962; DB 4; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-110; Mismatches 0; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; Patents 0; Sequence 3, Application US/09315689  
 ; Patent No. 6346510  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Philip E. Thorpe  
 ; APPLICANT: Rolf A. Breitken  
 ; TITLE OF INVENTION: ANTI BODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
 ; FILE REFERENCE: 4001-002584  
 ; CURRENT APPLICATION NUMBER: US/09/561,108  
 ; CURRENT FILING DATE: 2000-04-28  
 ; PRIOR FILING DATE: 1999-04-28  
 ; NUMBER OF SEQ ID NOS: 44  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 14  
 ; LENGTH: 182  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
 OTHER INFORMATION: PEPTIDE  
 US-09-561-108-14

Query Match 99.5%; Score 962; DB 4; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-110; Mismatches 0; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; Patents 0; Sequence 3, Application US/09315689  
 ; Patent No. 6346510  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Folkman, Judah  
 ; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions  
 ; FILE REFERENCE: 0513-0229  
 ; CURRENT APPLICATION NUMBER: US/09/315,689  
 ; CURRENT FILING DATE: 1999-05-20  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 178  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-315-689-5

Query Match 96.9%; Score 937; DB 4; Length 178;

Best local Similarity 100.0%, Pred. No. 9 1e-17; Mismatches 0; Indels 0; Gaps 0; Matches 178; Conservative 0;

Query 5 DRCPVVLVALNSPLSGMRCIRGADFOCFOQARAVGLAGTFRRAFLSSRLQDLYSIYRRA 64  
Db 1 DRCPVVLVALNSPLSGMRCIRGADFOCFOQARAVGLAGTFRRAFLSSRLQDLYSIYRRA 60

RESULT 6  
US-09-561-500-13  
Sequence 13, Application US/09561500  
Patent No. 6342219

GENERAL INFORMATION:  
APPLICANT: Philip E Thorpe  
APPLICANT: Rolf A. Brekken  
TITLE OF INVENTION: ANTI-BODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF FILE REFERENCE: 4001-002500  
CURRENT APPLICATION NUMBER: US/09/561,500  
CURRENT FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/131,432  
PRIOR FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Artificial Sequence

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-500-13

Query Match 86.9%; Score 840; DB 4; Length 191;  
Best Local Similarity 85.6%; Pred. No. 7.5e-95; Mismatches 15; Indels 0; Gaps 0; Matches 155; Conservative 15; MisMatches 11; Length: 191

Query 1 HSHRDFOQPVVLVALNSPLSGMRCIRGADFOCFOQARAVGLAGTFRRAFLSSRLQDLYSI 60  
Db 8 HTHQDFQPVVLVALNSPLSGMRCIRGADFOCFOQARAVGLAGTFRRAFLSSRLQDLYSI 67

Query 61 VRRADRAAVPIVNLKDELLPFWALFSGSEGPKPGARIFSDGKDVLRLRHPWPKSW 120  
Db 68 VRRADRGSPVIVNLKDEVLSPSWDSLFSQGQLQPGARIFSFDFGDVLRLRHPWPKSW 127

Query 121 HGSDPNGRLTESYCETWRTEAPSATQASSLIGGRIGQSAASCHAYIVLCIENSFT 180  
Db 128 HGSDPGRMLMESYCETWRTEAPSATQASSLISGRLEQKAASCHNSYIVLCIENSFT 187

Query 181 A 181  
Db 188 S 188

RESULT 8  
US-08-159-784-2  
Sequence 2, Application US/08159784  
Patent No. 5643783

GENERAL INFORMATION:  
APPLICANT: Bjorn R. Olsen  
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PC/2 Model 50Z or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159 784  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:  
NAME: John F. Freeman  
REGISTRATION NUMBER: 29, 066  
REFERENCE/DOCKET NUMBER: 00246/170001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELE: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 195  
TYPE: amino acid  
STRANDEDNESS: N/A

RESULT 7  
US-09-561-108-13  
Sequence 13, Application US/09561108  
Patent No. 6342219

GENERAL INFORMATION:  
APPLICANT: Philip E Thorpe  
TITLE OF INVENTION: ANTI BODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF FILE REFERENCE: 4001-002584  
CURRENT APPLICATION NUMBER: US/09/561,108  
CURRENT FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/131,432

PRIOR FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Artificial Sequence

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-108-13

Query Match 86.9%; Score 840; DB 4; Length 191;  
Best Local Similarity 85.6%; Pred. No. 7.5e-95; Mismatches 15; Indels 0; Gaps 0; Matches 155; Conservative 15; MisMatches 11; Length: 191

Query 1 HSHRDFOQPVVLVALNSPLSGMRCIRGADFOCFOQARAVGLAGTFRRAFLSSRLQDLYSI 60  
Db 8 HTHQDFQPVVLVALNSPLSGMRCIRGADFOCFOQARAVGLAGTFRRAFLSSRLQDLYSI 67

Query 61 VRRADRAAVPIVNLKDELLPFWALFSGSEGPKPGARIFSDGKDVLRLRHPWPKSW 120  
Db 68 VRRADRGSPVIVNLKDEVLSPSWDSLFSQGQLQPGARIFSFDFGDVLRLRHPWPKSW 127

Query 121 HGSDPNGRLTESYCETWRTEAPSATQASSLIGGRIGQSAASCHAYIVLCIENSFT 180  
Db 128 HGSDPGRMLMESYCETWRTEAPSATQASSLISGRLEQKAASCHNSYIVLCIENSFT 187

Query 181 A 181  
Db 188 S 188

TOPOLOGY: N/A  
US-08-159-784-2

Query Match 86.6%; Score 837; DB 1; Length 195;  
Best Local Similarity 85.1%; Pred. No. 1 8e-94;  
Matches 154; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

Qy 1 HSRDFQPVHLVALNPLSGMRRGADFOCFQQARAVGLAGTGRAFLSSRLQDLYSTI 60  
Db 12 HTHQDFQPVHLVALNPLSGMRRGADFOCFQQARAVGLAGTGRAFLSSRLQDLYSI 71

Qy 61 VRRADRAAVPTVNLKDELLFSPWEALFGSGEPLKPGARIFSFDFGKDVLRHTWPKSY 120  
Db 72 VRRADRSVPVNLDLDEVLSPLSWDSLFSGSQGOLQPGARIFSFDFGRDVRHPPQPSV 131

Qy 121 HGSDPNSRRLTSSYCEWRTEARSAQSLLGRLGQSASCHAYIVLCIENSFT 180  
Db 132 HGSDPSGRMLMESYCETWRTEGATGQASSLLSGRLLEORAASCHDSYIVLCIENSFT 191

Qy 181 A 181  
Db 192 S 192

RESULT 9  
US-08-985-526-36

; Sequence 36, Application US/08985526  
; Patent No. 6080728

GENERAL INFORMATION:

APPLICANT: MIXSON, James A  
TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA  
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
TITLE OF INVENTION: THERAPY  
NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: Connolly, Bove, Lodge, & Ruiz  
STREET: 1220 Market Street, P.O. Box 2207  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 19899

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPILER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/985, 526  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: McMorrow Jr., Robert G  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302) 658-9141  
TELEFAX: (302) 658-5613

INFORMATION FOR SEO ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

INFORMATION FOR SEO ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

US-08-985-526-36

Query Match 82.3%; Score 796; DB 3; Length 185;  
Best Local Similarity 82.4%; Pred. No. 1 7e-89;  
Matches 150; Conservative 16; Mismatches 14; Indels 2; Gaps 2;

Qy 1 HSRDFQPVHLVALNPLSGMRRGADFOCFQQARAVGLAGTGRAFLSSRLQDLYSI 60

Qy 61 VRRADRAAVPTVNLKDELLFSPWEALFGSGEPLKPGARIFSFDFGKDVLRHTWPKSY 119  
Db 61 VRRADRSVPVNLDLDEVLSPLSWDSLFSGSQGOLQPGARIFSFDFGRDVRHPPQPSV 120

Qy 120 WGSDPNSRRLTSSYCEWRTEARSAQSLLGRLGQSASCHAYIVLCIENSFT 179  
Db 121 WGSDPSGRMLMESYCETWRTEGATGQASSLLSGRLLEORAASCHDSYIVLCIENSFT 180

Qy 180 TA 181  
Db 181 TS 182

RESULT 10  
US-08-159-784-3

; Sequence 3, Application US/08159784  
; Patent No. 5643783

GENERAL INFORMATION:

APPLICANT: Bjorn R. Olsen  
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPILER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/159, 784  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: John F. Freeman  
REGISTRATION NUMBER: 229, 066  
REFERENCE/DOCKET NUMBER: 00246/170001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070  
TELEX: 200154  
TELEFAX: (617) 542-8906

INFORMATION FOR SEO ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 191  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: N/A

US-08-159-784-3

Query Match 53.4%; Score 516; DB 1; Length 191;  
Best Local Similarity 55.4%; Pred. No. 3 3e-55;  
Matches 98; Conservative 30; Mismatches 45; Indels 4; Gaps 1;

Qy 7 QPVHLVALNPLSGMRRGADFOCFQQARAVGLAGTGRAFLSSRLQDLYSTVRADE 66  
Db 18 KPALHLAALNNMPFGDIR---ADFOCFKQARAAGLSTRAPLSSHQLQDLYSTVRAE 73

Qy 67 AAIVPTVNLKDELLFSPWEALFGSGEPLKPGARIFSFDFGKDVLRHTWPKSYHGSDPN 126  
Db 74 YSLPTINLKGQVLNFNNWDSITSGHGGQPNMHPITPSDFGRDIMTDPSWPKVIVHGSPH 133

Qy 127 GRRLTSSYCEWRTEARSAQSLLGRLGQSASCHAYIVLCIENSFT 183

Db 134 GYVLVDNYCEAWFTADAVTGLASPLSNGKILQKAVSCANRLIVLTIENSPTMDAR 190

RESULT 11

US-09-046-985-2

; Sequence 2, Application US/09046985

; Patent No. 6121236

GENERAL INFORMATION:

; APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

TITLE OF INVENTION: ANGIOGENESIS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brock, Smith &amp; Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/474,743

FILING DATE:

CLASSIFICATION:

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 09/046,985

GENERAL INFORMATION:

; ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781) 861-6240

TELEFAX: (781) 861-9540

TELEPAK: (781) 861-9540

TELEMAIL: Patientin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781) 861-6240

TELEFAX: (781) 861-9540

TELEMAIL: Patientin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781) 861-6240

TELEFAX: (781) 861-9540

TELEMAIL: Patientin Release #1.0, Version #1.30

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/474,743  
FILING DATE:  
CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 09/046,985  
GENERAL INFORMATION:  
; ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781) 861-6240  
TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-474-743-2

Query Match 16.5%; Score 160; DB 3; Length 35;  
Best Local Similarity 100.0%; Pred. No. 9.9e-13; Mismatches 0; Indels 0; Gaps 0;

QY 36 QARAVGLAGTFRALSSLRQLDLYSIVRRADRAAV 69  
Db 1 QARAVGLAGTFRALSSLRQLDLYSIVRRADRAAV 34

RESULT 13  
US-09-046-985-7  
; Sequence 7, Application US/09046985  
Patent No. 6121236

GENERAL INFORMATION:

; APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

TITLE OF INVENTION: ANGIOGENESIS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brock, Smith &amp; Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781) 861-6240

TELEFAX: (781) 861-9540

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE: Modified-site

LOCATION: 1

OTHER INFORMATION: /label= "N-Acetyl-Threonine"

US-09-046-985-7

Query Match 10.4%; Score 101; DB 4; Length 22;

Best Local Similarity 100.0%; Pred. No. 8.4e-06;

Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Misnatches 0; Indels 0; Gaps 0;

45 TPRAFLSSRLQDLYSIVRRAD 65

Db 1 TERRAFLSSRLQDLYSIVRRAD 21

RESULT 14

US-09-474-743-7

Sequence 7, Application US/09474743

PATENT NO. 6235716

GENERAL INFORMATION:

APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPONENT: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/474,743

FILING DATE: 22-Oct-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Warren William L.

REGISTRATION NUMBER: 36,714

REFERENCE/DOCKET NUMBER: 05213-0223

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal.

ORIGINAL SOURCE:

ORGANISM: Murine

TISSUE TYPE: Collagen

US-08-740-168A-1

|||||||||||||||||

TPRAFLSSRLQDLYSIVRRAD 21

Sequence 1, Application US/08740168A

PATENT NO. 5854205

GENERAL INFORMATION:

APPLICANT: O'Reilly, Michael

APPLICANT: Folkman, M. Judah

TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew, LLP

STREET: 191 Peachtree, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPONENT: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/740,168A

FILING DATE: 22-Oct-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Warren William L.

REGISTRATION NUMBER: 36,714

REFERENCE/DOCKET NUMBER: 05213-0223

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal.

ORIGINAL SOURCE:

ORGANISM: Murine

TISSUE TYPE: Collagen

US-08-740-168A-1

|||||||||||||||||

HSHRDEQPVHLVALNSPLS 20

Sequence 1, Application US/08740168A

PATENT NO. 5854205

GENERAL INFORMATION:

APPLICANT: O'Reilly, Michael

APPLICANT: Folkman, M. Judah

TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew, LLP

STREET: 191 Peachtree, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPONENT: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/740,168A

FILING DATE: 22-Oct-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Warren William L.

REGISTRATION NUMBER: 36,714

REFERENCE/DOCKET NUMBER: 05213-0223

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal.

ORIGINAL SOURCE:

ORGANISM: Murine

TISSUE TYPE: Collagen

US-08-740-168A-1

|||||||||||||||||

HSHRDEQPVHLVALNSPLS 20

Sequence 1, Application US/08740168A

PATENT NO. 5854205

GENERAL INFORMATION:

APPLICANT: O'Reilly, Michael

APPLICANT: Folkman, M. Judah

TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew, LLP

STREET: 191 Peachtree, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPONENT: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/740,168A

FILING DATE: 22-Oct-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Warren William L.

REGISTRATION NUMBER: 36,714

REFERENCE/DOCKET NUMBER: 05213-0223

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal.

ORIGINAL SOURCE:

ORGANISM: Murine

TISSUE TYPE: Collagen

US-08-740-168A-1

|||||||||||||||||

HSHRDEQPVHLVALNSPLS 20

Sequence 1, Application US/08740168A

PATENT NO. 5854205

GENERAL INFORMATION:

APPLICANT: O'Reilly, Michael

APPLICANT: Folkman, M. Judah

TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew, LLP

STREET: 191 Peachtree, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPONENT: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/740,168A

FILING DATE: 22-Oct-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Warren William L.

REGISTRATION NUMBER: 36,714

REFERENCE/DOCKET NUMBER: 05213-0223

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 20 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: internal.

ORIGINAL SOURCE:

ORGANISM: Murine

TISSUE TYPE: Collagen

US-08-740-168A-1

|||||||||||||||||

HSHRDEQPVHLVALNSPLS 20

Sequence 1, Application US/08740168A

PATENT NO. 5854205

GENERAL INFORMATION:

APPLICANT: O'Reilly, Michael

APPLICANT: Folkman, M. Judah

TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jones & Askew, LLP

STREET: 191 Peachtree, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPONENT: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/740,168A

FILING DATE: 22-Oct-1996

CLASSIFICATION: 536

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 12, 2002, 10:10:43 ; Search time 30.23 Seconds

(without alignments)  
672.395 Million cell updates/secTitle: US-10-080-797-1  
Perfect score: 967  
Sequence: 1 HSHRDQPVVLVALNSPLS.....SCHAYIVLCIENSFTASK 183Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_032802:\*

- 1: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*
- 2: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*
- 3: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*
- 4: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*
- 5: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*
- 6: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*
- 7: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*
- 8: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*
- 9: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*
- 10: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*
- 11: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*
- 12: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*
- 13: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*
- 14: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*
- 15: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*
- 16: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*
- 17: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*
- 18: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*
- 19: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*
- 20: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*
- 21: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*
- 22: /\$IDS1/gcadata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result NO.	Score	Query Match Length	DB ID	Description
1	967	100.0	183 20	AY08693 Human endostatin P
2	967	100.0	183 20	AY02113 SEQ ID 76 of W991
3	967	100.0	183 21	AAB30493 Amino acid sequenc
4	967	100.0	183 21	AAB16451 Human endostatin P
5	967	100.0	183 21	AY90711 Human angiogenesis
6	967	100.0	183 21	AYT0252 Human angiogenes
7	967	100.0	183 22	AU00896 Human Endostatin(R)
8	967	100.0	183 22	AAB49379 Human endostatin S
9	967	100.0	216 21	AAB30495 Amino acid sequenc
10	967	100.0	684 18	RAW26327 Human alpha-1 coll
11	967	100.0	684 20	AYA25113 Human alpha1 (XVII

Result NO.	Score	Query Match Length	DB ID	Description	ALIGNMENTS
				RESULT 1	
				ID AY08693 standard; Protein: 183 AA.	
				ID AY08693; DT 10-AUG-1999 (first entry)	
				XX DE Human endostatin protein fragment.	
				XX KW Plasminogen; human; angiostatin; endostatin; gene therapy; vector; anti-angiogenic; attenuation; cytotoxic; anti-diabetic; ophthalmology; tumour growth; solid tumour; diabetic retinopathy; retina.	
				XX OS Homo sapiens.	
				XX PN WO9926480-A1.	
				PD 03-JUN-1999.	
				XX PF 20-NOV-1998; 98WO-US24950.	
				PR 20-NOV-1997; 97US-0975424.	
				PA (GENE-) GENETIX PHARM INC.	
				PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.	
				XX PT Bachelot T, Laboulch P, Pawlik RJ;	
				XX DR WPI: 1999-357696/30.	
				DR N-PSDB: AAX77719.	
				PT Anti angiogenic gene therapy vectors	

PS Disclosure: Page 74-75: 83pp: English

Bischoffsre: Page

This invention describes a novel viral gene therapy vector comprising a nucleic acid molecule encoding an anti-angiogenic polypeptide chosen from human or murine angiostatin, human or murine endostatin and angiogenesis-inhibiting fusions and fragments, where the viral vector is sufficiently attenuated for use in human gene therapy. The products of the invention have anti-angiogenic, cytostatic, anti-diabetic and ophthalmological activity. The vector is used in gene therapy for inhibiting tumour growth in humans harbouring a solid tumour. The vector expresses an anti-angiogenic polypeptide. An additional use comprises treatment of diabetic retinopathy, where the anti-angiogenic polypeptide inhibits angiogenesis in the vicinity of the retina. The vector is administered to cells ex vivo and then administered to the patient.

```

Query Match          100.0%; Score 967; DB 20; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.8e-110;
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

XX  
SO sequence 183 AA

Qy	1 HSIRDFOPVHLVALNSPLSGNNGRGAIDPOFCQDARAVLAGLFRRAFISRSRDLYSTI
Db	1 hsirdfopvhvalnsplsgnngraigdqcqfraravlaglfrrafisrsrdlysi 60
Qy	61 VRRADRAAVPTWNLKDELFLPSWEALFSGSGPLKGARIFSFQDGKDVLHPTWQKSW 120
Db	61 vrradraavptwnlkdelflpswealftsgsgplkgarisfdgpdvltptwqksw 120
Qy	121 HGSDPONGRLTESYCEWRTERAPSANTQASSLGGGLGQSASCHAYVLCIENSFT 180
Db	121 hgspdpongrltesycewtreapsatqgassllggqsaaschhayvlciensft 180
Qy	181 ASK 183
Db	181 ask 183

	Query Match	100.0%	Score	967	DB	20	Length	183;
	Best Local Similarity	100.0%	Pred. No.	9.8e-110				
	Best	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
Qy	HSHRFQPVHLJVALNSPLSGGMRGIRGAQDFCQFOARAVGLAGTFRATLSSRLQDLYSI	60						
1	H     S     R     F     Q     P     V     H     J     V     A     N     S     P     L     S     G     G     M     R     G     I     R     G     A     Q     D     F     C     Q     F     O     A     R     A     V     G     L     A     G     T     F     R     A     T     L     S     S     R     Q     D     L     Y     S     I	60						
Db	hshrdgqvphlvalnsplsggmrqrgaqdfqcfqgaravqagtaqtfralssrlqdyisi	60						
Qy	VRRADRAAPVITNLKDLFLPSWEALFGSEGSPLKPGARFSFDCKDVLEHPMPQKSW	120						
61	V     R     R     A     D     R     A     A     P     V     I     T     N     L     K     D     F     L     P     S     W     E     A     L     F     G     S     E     G     P     L     K     P     G     A     R     F     S     F     D     C     K     D     V     L     E     H     P     M     P     Q     K     S     W	120						
Db	vraddravapivnkkelffpswealfsgqegpkgrarifsfqgkdvirhpwqpksw	120						
Qy	121 HESDPNGRLRTESYCTEWRTEAPSATGQASSLGLGRLLGOSAASCHAYIVLCTENSMT	180						
121	H     E     S     D     P     N     G     R     L     R     T     E     S     Y     C     T     E     W     R     T     E     A     P     S     A     T     G     Q     A     S     S     L     G     L     G     O     S     A     S     C     H     A     Y     I     V     L     C     T     E     N     S     M     T	180						
Db	hgdpngrrltcsyctewrteapsaqgassllgrlgsgaaachhayivlcensmt	180						

```

RESULT 2
AAY02113
ID AAY02113 standard; Protein; 183 AA.
XX
AC AAY02113;
XX
DT 16-JUL-1999 (first entry)
XX
DE SEQ ID 76 of WO9916889.
XX
KW Antikostatin; endostatin; interferon; thrombospondin;
KW interferon-inducible protein; platelet factor 4; anti-angiogenic;
KW anti-tumor; multifunctional protein; angiogenic-mediated disease;
KW cancer; diabetic retinopathy; macular degeneration; arthritis;
KW tumor cell production.
XX
OS Homo sapiens.
XX
PN WO9916889-A1.
XX
PD 08-APR-1999.
XX
PF 30-SEP-1998; 98WO-US20464.
XX
PR 01-OCT-1997; 97US-0060609.
XX
PA (SEAR ) SEARLE & CO G D.
XX
PI Bolanowski MA, Caparon MH, Casperson GF, Gregory SA;
PT Klein BK, McKearn JP;
XX
DR WPI; 1999-255098/21.
XX
RP New multifunctional proteins useful for treating angiogenic-mediated

```

QY	181	ASK	183
Db	181	ask	183
<b>RESULT</b>	<b>3</b>		
AAB30493			
ID			
XX			
AAB30493	<b>standard;</b>	<b>Protein;</b>	<b>183 AA.</b>
AC			
XX			
DT	06-MAR-2001	(first entry)	
XX			
DE	Amino acid sequence of human endostatin encoded by plasmid pMALCH#15.		
XX			
KW	Streptomyces sp. strain C5; SpmA; S. venezuelae; alpha-amylase;		
KW	endostatin; cancer; tumour growth; angiogenesis.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2000050945-A1.		
XX			
PD	19-OCT-2000.		
XX			
PF	12-APR-2000; 2000WO-US09747.		
XX			
PR	13-APR-1999; 99US-0129084.		
XX			
PA	(MERI ) MERCK & CO INC.		
XX			
PI	Desanti CL, Strohl WR;		
XX			
DR	WPI: 2000-686970-67.		
DR	N-PSDB; AACG2023.		
XX			

PT Preparation of soluble recombinant endostatin involves transforming  
 PT Streptomyce host with expression vector comprising nucleotide  
 PT sequence encoding endostatin operably linked to linker and leader  
 PT peptide -  
 XX  
 PS Example 1; FIG 6; 57pp; English.

XX The present sequence represents human endostatin. The protein is  
 CC expressed in Streptomyces. Leader sequences of Streptomyces sp. strain  
 CC C5 SnpA and S. venezuelae alpha-amylase proteins are linked to the  
 CC N-terminal of endostatin. This ensures that endostatin protein is  
 CC produced as a secreted, soluble protein which needs no refolding, is  
 CC stable in the fermentation broth and is produced in large quantities.  
 CC The method is used for preparing soluble recombinant human, murine or  
 CC primate endostatin, which is useful in the treatment of cancer,  
 CC inhibition of tumour growth, inhibition of angiogenesis, isolation of  
 CC receptors for endostatin and for identification of anti-angiogenic  
 CC compounds in assays. The endostatin protein is produced as a secreted,  
 CC soluble protein which needs no refolding, is stable in the fermentation  
 CC broth and is produced in large quantities. Streptomyces are amenable  
 CC for cultivation in large fermentations allowing for large quantities of  
 XX soluble endostatin to be produced.

SQ Sequence 183 AA:

Query Match 100.0%; Score 967; DB 21; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-110; Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGGMRGIRGAQDFQOQARAVLAGTFRALFSSRLQDLYSI 60  
 1 hshrdqpvhlvalnsplsggmrqrgadfqcfqgavrlqgatfralflssrlqdlysi 60

QY 61 VRRADRAAVPIVNLKDELLFSPSWEALFSGSEGKLGPKGARIFSDGKDVLRHPWPKSW 120  
 61 vrradraavpivnlkdeellfspswalfsqsgpkpgarifsfqgdkvrlrhpwpqksvw 120

QY 121 HGSPDPNGRRLTESCYETWRTEASATGQASSLGLGRLGQSAA SCHHAYIVLICENSMT 180  
 121 hgdpngrltescyetwrteasatgqassllgrlgqsaaschhayivlicensmt 180

QY 181 ASK 183  
 181 ask 183

Db 181 ask 183

RESULT 4

AAB16451 standard; Protein; 183 AA.

XX AAB16451:  
 AC  
 XX DT 27-OCT-2000 (first entry)  
 XX DE Human endostatin protein sequence.  
 XX  
 KW Andiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;  
 KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;  
 KW psoriasis; scleroderma; myocardial angiogenesis; Crohn's Disease;  
 KW cerebral collateral; arteriovenous malformation; rubrosis; cancer;  
 KW diabetic retinopathy; arthritis; wound healing; peptic ulcer;  
 KW Helicobacter related disease; fracture; cat scratch fever;  
 XX Homo sapiens.  
 XX WO200032631-A2.

XX 08-JUN-2000.  
 XX 06-DEC-1999; 99NO-US28897.  
 XX 04-DEC-1998; 9805-0206059.

XX  
 PA (ENTR-) ENTREMED INC.  
 PT XX  
 PT Macdonald NJ, Sim KL;  
 PI XX  
 PS DR WPI: 2000-412290/35.

XX New angiogenesis-inhibiting protein receptors, useful in methods for  
 PT treating diseases and processes that are mediated by angiogenesis, such  
 PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -  
 XX Disclosure: Figure 3; 10pp; English.

CC This invention relates to angiogenesis-inhibiting protein receptors, and  
 CC the DNA sequences encoding them. Angiogenesis is the generation of new  
 CC blood vessels into a tissue, and normally occurs in 'wound healing', 'luteum',  
 CC 'foetal' and 'embryonal' development and the formation of the corpus luteum,  
 CC endometrium and placenta. Endostatin is a protein (see AAB16450 and  
 CC AAB16202) involved in angiogenesis, and has an amino acid sequence  
 CC similar to that of a plasminogen fragment (see murine plasminogen  
 CC AAB16490). Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and  
 CC AAB16203). Sequences AAB16424 and AAB16522 represent coding and protein  
 CC sequences of human laminin. Laminin is an angiostatin binding protein,  
 CC and some of the peptides of the invention share homology with regions of  
 CC laminin. Peptides AAB1652-B16521 (excluding AAB16490) are the  
 CC peptides bind either angiostatin or endostatin and can be used in methods  
 CC for treating diseases and processes that are mediated by angiogenesis,  
 CC such as solid tumours, psoriasis, Crohn's disease, cerebral collaterals, arteriovenous malformations,  
 CC rubrosis, diabetic retinopathy, fractures, arthritis, wound healing, peptic ulcers,  
 CC Helicobacter related diseases, placenta, fever. They are useful for the detection and prognosis of cancer. DNA  
 CC sequences A628204 A628241 encode the peptides of the invention.

SQ Sequence 183 AA;

Query Match 100.0%; Score 967; DB 21; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-110; Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGGMRGIRGAQDFQOQARAVLAGTFRALFSSRLQDLYSI 60  
 1 hshrdqpvhlvalnsplsggmrqrgadfqcfqgavrlqgatfralflssrlqdlysi 60

QY 61 VRRADRAAVPIVNLKDELLFSPSWEALFSGSEGKLGPKGARIFSDGKDVLRHPWPKSW 120  
 61 vrradraavpivnlkdeellfspswalfsqsgpkpgarifsfqgdkvrlrhpwpqksvw 120

QY 121 HGSPDPNGRRLTESCYETWRTEASATGQASSLGLGRLGQSAA SCHHAYIVLICENSMT 180  
 121 hgdpngrltescyetwrteasatgqassllgrlgqsaaschhayivlicensmt 180

QY 181 ASK 183  
 181 ask 183

Db 181 ask 183

RESULT 5

AY90771 standard; Protein; 183 AA.

ID AAY90771:  
 AC AAY90771;  
 XX DT 22-AUG-2000 (first entry)  
 XX DE Human angiogenesis inhibiting factor 1 protein.  
 XX KW Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;  
 KW abnormal vessel disease.

XX OS Homo sapiens.  
 XX CN1244536-A.  
 PN XX  
 PN WO2001133-A2.  
 PD XX  
 PD 02-MAR-2000.  
 PR XX  
 PR 25-AUG-1998; 99WO-US19329.  
 XX PR 25-AUG-1998; 98US-0097883.  
 PA XX  
 PA (LEXIT-) LEXINGEN PHARM CORP.  
 PI XX  
 PI Lo K., Li Y., Gillies SD;  
 XX DR WPI; 2000-237616/20.  
 DR N-PSDB; AA251291.  
 XX  
 PT Angiogenesis inhibiting factor 1 and its derivative useful for treating  
 tumors -  
 XX SQ Sequence 183 AA:  
 PS Claim 1; Fig 5; 41pp; Chinese.  
 CC The present sequence represents an angiogenesis inhibiting factor (1),  
 designated IAF-1. The present invention also describes: (1) preparation  
 of (1) and its derivative; (2) an IAF binding acceptor and its  
 preparation; and (3) an IAF antibody. (1) is useful for preparing new  
 biological preparations for effectively treating various tumours and  
 abnormal-vessel diseases. The IAF antibody is preferably a polyclonal  
 antibody, mosaic antibody, single stranded antibody and human originated  
 antibody.  
 XX SQ Sequence 183 AA:  
 CC Query Match 100.0%; Score 967; DB 21; Length 183;  
 CC Best Local Similarity 100.0%; Pred. No. 9.8e-110; Mismatches 0; Indels 0; Gaps 0;  
 CC Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 QY 1 HSHRDFQPVHLVALNNSPLSGGMRGIRGADFQCFQARAVGLAGTFRALFLSSRLQDLYSI 60  
 CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 CC 1 hshrdfqpvhlvalnsplsggmgirgadfqcfqaravglagtfraflslrqdlysi 60  
 CC  
 QY 61 VRRADRAAVPVINVKDELFLPSWEALNSPLSGSEGPKPGARIFSPDGKDVLRLHPTWPKSW 120  
 CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 CC 61 vrradraavpvinvkdelflpswealtsplsggpkpgarifspdgkdvlrlhptwpgksw 120  
 CC  
 QY 121 HGSDPNGRRLTESCYCEWRTEAPSATGQASSLGLGQASANSCHHAYVLCIENSMT 180  
 CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 CC 121 hgsdpngrrltescyctwrteapsatggassllgqrlqsaaschhayivciensmt 180  
 CC  
 Db  
 QY 181 ASK 183  
 Db 181 ask 183  
 RESULT 6  
 AY70252 ID AY70252 standard; Protein; 183 AA.  
 AC AY70252:  
 XX  
 DT 06-JUN-2000 (first entry)  
 DE Human angiogenesis inhibitor, endostatin.  
 KW Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;  
 KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antilarthritic;  
 KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;  
 KW vasoconstrictive; treatment; antiarteriosclerosis; tumour;  
 KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
 KW wound granulation; keloid scar; gene therapy.

---

XX OS Homo sapiens.  
 XX CN1244536-A.  
 PN XX  
 PN WO2001133-A2.  
 PD XX  
 PD 02-MAR-2000.  
 PR XX  
 PR 25-AUG-1998; 99WO-US19329.  
 XX PR 25-AUG-1998; 98US-0097883.  
 PA XX  
 PA (LEXIT-) LEXINGEN PHARM CORP.  
 PI XX  
 PI Lo K., Li Y., Gillies SD;  
 XX DR WPI; 2000-237616/20.  
 DR N-PSDB; AA251291.  
 XX  
 PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin  
 PT FC region, useful for treating conditions mediated by angiogenesis,  
 PT such as rheumatoid arthritis, tumors and macular degeneration -  
 XX SQ Sequence 183 AA:  
 PS Example 1; Pages 41-42; 68pp; English.  
 CC The patent discloses a DNA molecule encoding a fusion protein comprising  
 CC a signal sequence, an immunoglobulin FC region, and an angiogenesis  
 CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment  
 CC having angiostatin activity, a collagen XVIII fragment having endostatin  
 CC activity, or combinations of them. The fusion protein (immunofusin) is  
 CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
 CC by angiogenesis. Conditions that may be treated include solid tumours,  
 CC blood born tumours, tumour metastasis, benign tumours including  
 CC haemangiomas, acoustic neuroma, neurofibromas, trachoma and pyrogenic  
 CC granulomas, rheumatoid arthritis, psoriasis, oculal angiogenic diseases  
 CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
 CC degeneration, corneal graft rejection, neovascular glaucoma, retrothalamic  
 CC fibroplasia, rubeosis and Osler-Weber syndrome; myocardial angiogenesis,  
 CC plaque neovascularisation, telangiectasia, haemodialysis joints.  
 CC angiofibroma, wound granulation, ectasia, haemorrhage or abnormal joint.  
 CC endochelial cells, intestinal cells, atherosclerosis, sclerodermal and  
 CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
 CC in gene therapy. The present sequence is a human endostatin used in the  
 CC construction of immunofusin containing human immunoglobulin gamma  
 CC (IgG) FC fragment.  
 XX SQ Sequence 183 AA:  
 CC Query Match 100.0%; Score 967; DB 21; Length 183;  
 CC Best Local Similarity 100.0%; Pred. No. 9.8e-110; Mismatches 0; Indels 0; Gaps 0;  
 CC Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 QY 1 HSHRDFQPVHLVALNNSPLSGGMRGIRGADFQCFQARAVGLAGTFRALFLSSRLQDLYSI 60  
 CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 CC 1 hshrdfqpvhlvalnsplsggmgirgadfqcfqaravglagtfraflslrqdlysi 60  
 CC  
 QY 61 VRRADRAAVPVINVKDELFLPSWEALNSPLSGSEGPKPGARIFSPDGKDVLRLHPTWPKSW 120  
 CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 CC 61 vrradraavpvinvkdelflpswealtsplsggpkpgarifspdgkdvlrlhptwpgksw 120  
 CC  
 QY 121 HGSDPNGRRLTESCYCEWRTEAPSATGQASSLGLGQASANSCHHAYVLCIENSMT 180  
 CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 CC 121 hgsdpngrrltescyctwrteapsatggassllgqrlqsaaschhayivciensmt 180  
 CC  
 Db  
 QY 181 ASK 183  
 Db 181 ask 183  
 RESULT 7  
 ID AAU00896  
 ID AAU00896 standard; Protein; 183 AA.  
 XX

AC	AAU00896;	Db	1 hshdrfqpvihlvaainspqsgmrgqjrgadfqgqaravglagttraflissrlQdlysi
DT	04 -JUL-2001 (first entry)	QY	61 VRRADRAAVIVLNKDELIFPSWEALFSGSGPQPKGARIFSFDGKVLRHPTWPKSVW 120
DE	Human Endostatin(TM) protein.	Db	61 vrradaavivlnkdelifpswealfsgsgpqlparifsfdfgadvrlptwpksvw 120
XX	Human; Endostatin(TM); angiogenesis mediated disease; solid tumours; blood borne tumour; leukaemia; tumour metastasis; benign tumour; haemangioma; acoustic neuroma; neurofibroma; trachoma; rubesis; pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer; ocular angiogenic disease; diabetic retinopathy; macular degeneration; retinopathy of prematurity; macular corneal graft rejection; neovascular glaucoma; retrorenal fibroplasia; Osler-Weber Syndrome; myocardial angiogenesis; plaque neovascularisation; telangiectasia; haemophilic joint; angiofibroma; wound granulation.	QY	121 HGSDDNGRRLTESCETWRTEBAPSANGQASSLIGGRGQASASCHAYAVTLIENSFT 180
KW		Db	121 hgspngritesyctewrteapsatqgassllgqsaaschayivlciensft 180
KW		QY	181 ASK 183
KW		Db	181 ask 183
OS	Homo sapiens.	RESULT	8
XX	PN WO200119389-A2.	ID	AAB49379
XX	PD 22-MAR-2001.	ID	AAB49379 standard; Protein; 183 AA.
XX	PF 14-SEP-2000; 2000WO-US25166.	ID	AAB49379;
XX	PR 14-SEP-1999; 99US-015698.	ID	AAB49379;
XX	PA (ENTR-) ENTREMED INC.	DE	02-MAR-2001 (first entry)
XX	PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ; Bermejo LI, Mistry FR, Shepard SR, Schrimsher JL;	DT	Human endostatin SEQ ID NO: 2.
XX	DR DR N-PSDB; AAS00867.	XX	Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken; cancer; inflammation; angiogenesis-dependent disease.
XX	PT Producing Endostatin protein for treating angiogenesis mediated diseases such as solid tumours, comprises recombinantly producing the protein -	XX	Homo sapiens.
PT	Producing Endostatin protein for recovering and purifying the protein using an expression system, and recovering and purifying the	XX	OS Homo sapiens.
PT	Producing Endostatin protein for recovering and purifying Endostatin (TM) from biological sources, such as biological fluids, tissues, cells, culture media, and fermentation media. Endostatin(TM) is useful for treating angiogenesis mediated diseases such as solid tumours, blood borne tumours, leukaemias, tumour metastases, benign tumours, e.g., haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases, e.g., diabetic retinopathy, retinopathy of prematurity, macular degeneration, corneal graft rejection, neovascular glaucoma, colon cancer, retrorenal fibroplasia, rubesis, Oster-Webber Syndrome, myocardial angiogenesis, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, and wound granulation. Endostatin(TM) is also useful for treating disease of excessive or abnormal stimulation of endothelial cells such as intestinal adhesions, atherosclerosis, scleroderma and hypertrophic scars. Higher yields of more purified, and biologically active Endostatin(TM) are obtained by the new method.	XX	WO20006771-A1.
XX	CC	XX	PN 16-NOV-2000.
XX	CC	XX	PR 02-MAY-2000; 2000WO-US12063.
XX	CC	XX	PR 06-MAY-1999; 99US-0132907.
XX	CC	XX	PR 14-JUL-1999; 99US-0353333.
XX	CC	XX	PA (BURN-) BURNHAM INST.
XX	CC	XX	PI Vuori K;
XX	CC	XX	PR WPI; 2001-040937/05.
XX	CC	XX	DR DR N-PSDB; AAC88289.
XX	CC	XX	PT Endostatin peptide comprising at least four endostatin amino acid residues are e.g. angiogenesis inhibitors for treating cancer and diabetic retinopathy -
XX	CC	XX	PT Disclosure; Fig 1; 146pp; English.
XX	CC	XX	The present invention provides endostatin peptides which can be used in the modulation of angiogenesis. This is useful in the treatment of cancers, inflammation, rheumatoid arthritis, chronic articular rheumatism, psoriasis, disorders associated with inappropriate invasion of vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy of prematurity, macular degeneration, corneal graft rejection, retrorenal fibroplasia, rubesis, capillary proliferation in atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque neovascularisation, telangiectasia, haemophilic joints and wound granulation. In addition, the peptides can be used as birth control agents.
SQ	Sequence 183 AA;	CC	

```

Qy      1 HSHRDFOQPVHLVALNSPLSGGMRGIRGADEFQCFQOARANGLAGTFRALFSSLRQLDLYSI 60
      |          |          |          |          |          |          |          |
      Query Match 100.0%; Score 967; DB 22; Length 183;
      Best Local Similarity 100.0%; Pred. No. 9.8e-110;
      Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

Query 1 HSIRDQFQVPLHVALNSPLSLGGMRGRCAGADFCQFCQFQARAVGLACTRAFLSSRLQDLYSI 60  
 CC soluble protein which needs no refolding, is stable in the fermentation  
 CC broth and is produced in large quantities. Streptomyces are amenable  
 CC for cultivation in large fermentations allowing for large quantities of  
 CC soluble endostatin to be produced.

CC compounds in assays. The endostatin protein is produced as a secreted,  
 CC recombinant human endostatin in Streptomyces. Leader sequences of  
 CC Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins  
 CC are linked to the N-terminal of endostatin. This ensures that endostatin  
 CC protein is produced as a secreted, soluble protein which needs no  
 CC refolding, is stable in the fermentation broth and is produced in large  
 CC quantities. The method is used for preparing soluble recombinant human,  
 CC murine or primate endostatin, which is useful in the treatment of cancer,  
 CC inhibition of tumour growth, inhibition of angiogenesis, isolation of  
 CC receptors for endostatin and for identification of anti-angiogenic  
 CC receptors.

Db 1 hshrdfqpvihivalnsplsgmrgirgadfcfcqqravglagltfrafraflssrlqdlysi 60  
 Qy 61 VRRADRAAVPVLKDELLFPLSWEALFGSGEPLKPGARRIFSDGKDVLRHPTWPKSWV 120  
 Db 61 vrradraavpvlkdeffpswealfgsgcplkgarrifsdgkvdrhptwpgksvw 120  
 Qy 121 HGSDPNGRRLTESYCETWRTEAPSATQASSLIGGRILGOSASAACHAYIVLCIENSFMT 180  
 Db 121 hgsdpngrltesycetwrteapsatqgassllggrilgqsaaschayivlcienfmt 180  
 Qy 181 ASK 183  
 Db 181 ask 183  
 Result 9  
 AAB30495  
 ID AAB30495 standard; protein; 216 AA.  
 XX  
 AC AAB30495;  
 DT 06-MAR-2001 (first entry)  
 XX  
 DE Amino acid sequence of vaa-endostatin fusion protein in PANT3052.  
 XX  
 KW Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;  
 endostatin; cancer; tumour growth; angiogenesis.  
 XX  
 OS Synthetic.  
 OS Streptomyces sp.  
 OS Homo sapiens.  
 FH Location/Qualifiers  
 FT Peptide 1..28  
 /note= "vaa signal sequence"  
 FT Protein 29..216  
 /note= "endostatin"  
 PN WO200050945-A1.  
 PD 19-OCT-2000.  
 XX  
 PF 12-APR-2000; 2000WO-US09747.  
 XX  
 PR 13-APR-1999; 99US-0129084.  
 PA (MERRI ) MERCK & CO INC.  
 XX  
 PI Desantti CL, Strohl WR;  
 XX  
 DR WPI; 2000-586970/67.  
 DR N-PSDB; AAC62025.  
 XX  
 PT Preparation of soluble recombinant endostatin involves transforming  
 PT Streptomyces host with expression vector comprising nucleotide  
 PT sequence encoding endostatin operably linked to linker and leader  
 PT peptide -  
 PS Example 1; Fig 10A-B; 57pp; English.  
 XX  
 CC The present sequence represents a fusion protein of vaa and endostatin.  
 CC The specification language describes a method for the production of soluble,  
 CC recombinant human endostatin in Streptomyces. Leader sequences of  
 CC Streptomyces sp. strain C5 SnpA and S. venezuelae alpha-amylase proteins  
 CC are linked to the N-terminal of endostatin. This ensures that endostatin  
 CC protein is produced as a secreted, soluble protein which needs no  
 CC refolding, is stable in the fermentation broth and is produced in large  
 CC quantities. The method is used for preparing soluble recombinant human,  
 CC murine or primate endostatin, which is useful in the treatment of cancer,  
 CC inhibition of tumour growth, inhibition of angiogenesis, isolation of  
 CC receptors for endostatin and for identification of anti-angiogenic  
 CC receptors.

Query Match 100.0%; Score 967; DB 21; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 1 3e-109;  
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Db 1 HSIRDQFQVPLHVALNSPLSLGGMRGRCAGADFCQFCQFQARAVGLACTRAFLSSRLQDLYSI 60  
 Qy 61 VRRADRAAVPVLKDELLFPLSWEALFGSGEPLKPGARRIFSDGKDVLRHPTWPKSWV 120  
 Db 61 vrradraavpvlkdeffpswealfgsgcplkgarrifsdgkvdrhptwpgksvw 120  
 Qy 94 vrradraavpvlkdeffpswealfgsgcplkgarrifsdgkvdrhptwpgksvw 153  
 Db 94 vrradraavpvlkdeffpswealfgsgcplkgarrifsdgkvdrhptwpgksvw 153  
 AC AAM26327;  
 DT 19-NOV-1997 (first entry)  
 XX  
 DE Human alpha-1 collagen (XVIII).  
 XX  
 KW Alpha-1 collagen; type XVIII collagen; cartilage degeneration.  
 XX  
 OS Homo sapiens.  
 Result 10  
 AAM26327  
 ID AAM26327 standard; Protein; 684 AA.  
 XX  
 AC AAM26327;  
 DT 19-NOV-1997 (first entry)  
 XX  
 DE Human alpha-1 collagen (XVIII).  
 XX  
 KW Alpha-1 collagen; type XVIII collagen; cartilage degeneration.  
 XX  
 OS Homo sapiens.  
 FH Location/Qualifiers  
 FT Key 1..6  
 FT Peptide 1..6  
 /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 7..12  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 13..18  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 19..24  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 25..30  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 31..36  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 37..42  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 48..53  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 54..59  
 FT /label= GXGX'Y'\_motif  
 FT /note= "Claim 1"  
 FT Peptide 74..79



XX Novel human type alpha-1 (XVIII) collagen is characterised by  
 CC 10 triple helical domains containing the GXYGX<sub>n</sub>Y motif (where X,  
 CC Y, X' and Y' represent any amino acid), the helical domains being  
 CC separated and flanked by non triple helical regions which may  
 CC provide flexibility. Alpha-1 collagen is expressed in multiple  
 CC tissues, especially liver, lung and kidney. A cloned plasmid  
 CC comprising alpha-1 collagen nucleic acid (see Art8484) and an  
 CC expression control sequence can be used to express recombinant  
 CC collagen in prokaryotic or eukaryotic (especially mammalian) host  
 CC cells. The alpha-1 collagen may be used to treat a patient  
 CC suffering from a disease associated with cartilage degradation, and  
 CC for supplementing collagen. It can also be used as a connective  
 CC tissue.

Query Match 100.0%; Score 967; DB 18; Length 684;  
 Best Local Similarity 100.0%; Pred. No. 7e-109;  
 Matches 183; Conservatve 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HSIRDQPVHLVALNSPLSGGGRGIRGADRCFCQFQARAVLAGTTRAFRLSSRLDLYSI 60  
 Db 502 hshrdfqpvhlvalnsplsgggrgirgadrcfcqfqqaravlagttrafrlssrldlysi 561

Qy 61 VRADRAAVPTVNKLDELFPSNEALFGSGECPKPCARIFSDFGKVLRHTWPQKSW 120  
 Db 562 vraddravpvnkldeelfpsnealfsgsgcpkpcarifsdgkvdrhtwpqksw 621

Qy 121 HGSDPNGRRLTESYCEWRTEARSATGQASSLGLGRULGOSAASCHAYIVLCIENSFMT 180  
 Db 622 hgsdpngrrltesycetwrteapsatgqassllgrulgosaaschayivlcienfmt 681

Qy 181 ASK 183  
 Db 682 ask 684

RESULT 11

ID AAV25113 standard; Protein; 684 AA.  
 AC AAV25113;  
 XX  
 DT 25-AUG-1999 (first entry)  
 DE Human alpha1 (XVIII) collagen protein.  
 XX  
 KW Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;  
 KW anti-angiogenic; heparin binding domain; receptor binding domain; mimic;  
 KW alpha-helix A domain; carbohydrate recognition domain; CRD domain;  
 KW treatment; angiogenesis; tumour; human.  
 OS Homo sapiens.  
 PN W09931616-A1.  
 XX  
 PD 24-JUN-1999.  
 XX  
 PF 16-DEC-1998; 98WO-US26783.  
 PR 16-DEC-1997; 97US-0069727.

XX (HARD ) HARVARD COLLEGE.  
 XX  
 PT Hohenester E, Olsen BR, Sasaki T, Timpl R;  
 XX WPI; 1999-395243/33.  
 DR N-PSDB; AAXT8379.

XX Identifying mimetics of mammalian endostatin  
 PT  
 XX Disclosure; Fig 5A-C; 75pp; English.  
 PS XX  
 CC This invention describes a novel method for identifying mimetics of  
 mammalian endostatin. The method comprises identifying a compound

CC having atomic coordinates with non-trivial similarity to selected  
 CC coordinates of atoms of a mammalian endostatin involves (a) providing  
 CC a library of atomic coordinates of compounds in a library of candidate  
 CC compounds, (b) comparing the library of atomic coordinates to the  
 CC selected coordinates of a mammalian endostatin and (c) selecting from the  
 CC library at least one candidate compound on the basis of selection  
 CC criteria which include similarities between the atomic coordinates of the  
 CC selected candidate compound and the atomic coordinates of the mammalian  
 CC endostatin. The invention also describes the use of an anti-angiogenic  
 CC fragment of endostatin comprising a domain selected from a heparin  
 CC binding domain, a receptor binding domain, and exposed on alpha-helix A  
 CC domain, and a carbohydrate recognition domain (CRD) domain. The methods  
 CC can be used for designing and selecting endostatin mimics. The compounds  
 CC identified can be used for treating undesired angiogenesis, e.g. tumours.  
 CC This sequence represents human alpha1(XVIII) collagen which is used in  
 CC the description of the method.

Sequence 684 AA:

Query Match 100.0%; Score 967; DB 20; Length 684;  
 Best Local Similarity 100.0%; Pred. No. 7e-109;  
 Matches 183; Conservatve 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HSIRDQPVHLVALNSPLSGGGRGIRGADRCFCQFQARAVLAGTTRAFRLSSRLDLYSI 60  
 Db 502 hshrdfqpvhlvalnsplsgggrgirgadrcfcqfqqaravlagttrafrlssrldlysi 561

Qy 61 VRADRAAVPTVNKLDELFPSNEALFGSGECPKPCARIFSDFGKVLRHTWPQKSW 120  
 Db 562 vraddravpvnkldeelfpsnealfsgsgcpkpcarifsdgkvdrhtwpqksw 621

Qy 121 HGSDPNGRRLTESYCEWRTEARSATGQASSLGLGRULGOSAASCHAYIVLCIENSFMT 180  
 Db 622 hgsdpngrrltesycetwrteapsatgqassllgrulgosaaschayivlcienfmt 681

Qy 181 ASK 183  
 Db 682 ask 684

RESULT 12

ID AAW92296 standard; peptide; 1301 AA.  
 XX  
 AC AAW92296;  
 DT 28-APR-1999 (first entry)  
 DE Human alpha-1 (XVIII) collagen chain common sequence HJ18(common)36.  
 XX  
 KW Human; type XVIII collagen; liver disease; cirrhosis; detection;  
 XX hepatocellular carcinoma; diagnosis.  
 OS Homo sapiens.  
 XX  
 PN W09856399-A1.  
 XX  
 PD 17-DEC-1998.  
 XX  
 PF 12-JUN-1998; 98WO-US12327.  
 PR 12-JUN-1997; 97US-0049369.  
 XX  
 PA (FIFT-) ACAD FINLAND.  
 PA (FIBR-) FIBROGEN INC.  
 PA (INRM ) INSTNAT SANTE & RECH MEDICALE.  
 XX  
 PT Clement B, Pihlajaniemi T, Rehn M;  
 XX WPI; 1999-070292/06.  
 XX  
 Diagnose and monitoring of liver disease by measuring collagen type



XX Endostatin peptide comprising at least four endostatin amino acid  
 PT residues are e.g. angiogenesis inhibitors for treating cancer and  
 PT -diabetic retinopathy -  
 XX  
 PS Claim 17; Page 125-126; 146pp; English.

XX The present invention provides endostatin peptides which can be used in  
 CC the modulation of angiogenesis. This is useful in the treatment of  
 CC cancers, inflammation, rheumatoid arthritis, chronic articular  
 CC rheumatism, psoriasis, disorders associated with inappropriate invasion of  
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy  
 CC of prematurity, macular degeneration, corneal graft rejection, in  
 CC retrolental fibroplasia, rubesis, capillary proliferation in  
 CC diseases such as atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent  
 CC diseases include Osler-Webber syndrome, haemophilic joints and wound  
 CC granulation. In addition, the peptides can be used as birth control  
 CC agents.  
 CC

SQ Sequence 183 AA;

Query Match 99.6%; Score 963; DB 22; Length 183;  
 Best Local Similarity 99.5%; Pred. No. 3e-109; Mismatches 0; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; MisMatches 1; Del 0; Insert 0;

OY 1 HSIRDFOPLVHLVALNSPLSGGMRGIRGADFOCQQARAVLAGTFRAGLFLRSRQLDLYSTI 60  
 Db 1 hshrdfqpvihivalnsplsggmrgrgadfqcfqgaravlagtfragraffslridlysi 60

OY 61 VRRADRAAVPTVNLKDELLEPSWEALFSGSGCPKPGARTFSFDGKVLRHTPQKSWV 120  
 Db 61 viradravlpvnikdellepswealfsgsgcpkpgarifsfqdgkvrlhtpwpksvw 120

OY 121 HGSDPNGRRLTESYCERWRTEAPSATQASSLLGRGLGOSAASCHAYIVLCIENSFMT 180  
 Db 121 hgsdpngrrltesycetwrteapsatqassllgrglgosaaschayivlciensfmt 180

OY 181 ASK 183  
 Db 181 bgsdpngrrltesycetwrteapsatqassllgrglgosaaschayivlciensfmt 180

OY 181 ask 183  
 Db 181 ask 183

RESULT 15  
 AAB28399 AAB28399 standard; Protein; 182 AA.  
 XX  
 AC AAB28399;  
 XX  
 DT 19-FEB-2001 (first entry)  
 DE Human endostatin.  
 XX Human; endostatin; cytostatic; antiproliferative;  
 KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;  
 XX cancer; vascularised solid tumour.  
 OS Homo sapiens.  
 XX  
 PN WO20064946-A2.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 28-APR-2000; 2000WO-US11367.  
 XX  
 PR 28-APR-1999; 99US-0131432.  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PT Thorpe PE, Brekken RA;  
 XX  
 DR, WPI; 2000-687317/67.

Query Match 99.5%; Score 962; DB 21; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 4e-109; Mismatches 0; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; MisMatches 1; Del 0; Insert 0;

OY 1 HSIRDFOPLVHLVALNSPLSGGMRGIRGADFOCQQARAVLAGTFRAGLFLRSRQLDLYSTI 60  
 Db 1 hshrdfqpvihivalnsplsggmrgrgadfqcfqgaravlagtfragraffslridlysi 60

OY 61 VRRADRAAVPTVNLKDELLEPSWEALFSGSGCPKPGARTFSFDGKVLRHTPQKSWV 120  
 Db 61 viradravlpvnikdellepswealfsgsgcpkpgarifsfqdgkvrlhtpwpksvw 120

OY 121 HGSDPNGRRLTESYCERWRTEAPSATQASSLLGRGLGOSAASCHAYIVLCIENSFMT 180  
 Db 121 hgsdpngrrltesycetwrteapsatqassllgrglgosaaschayivlciensfmt 180

OY 181 AS 182  
 Db 181 as 182

Search completed: August 12, 2002, 10:13:06  
 Job time: 143 sec

XX Immunogenic composition for the treatment and diagnosis of cancer  
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody  
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -  
 XX  
 PS Example 10; Page 291-292; 298pp; English.

XX The present invention relates to anti-Vascular Endothelial Growth Factor  
 CC (VEGF) antibodies that bind to the same epitope as the monoclonal  
 CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to  
 CC the VEG receptor VEGFR, without inhibiting VEGF binding to the VEG  
 CC receptor VEGFR1. The present sequence is human endostatin. Endostatin  
 CC may be conjugated onto the anti-VEGF antibodies of the present invention.  
 CC The anti-VEGF antibodies of the present invention are useful for the  
 CC treatment and diagnosis of cancer, especially vascularised solid tumours.  
 XX

SQ Sequence 182 AA;

Query Match 99.5%; Score 962; DB 21; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 4e-109; Mismatches 0; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; MisMatches 1; Del 0; Insert 0;

OY 1 HSIRDFOPLVHLVALNSPLSGGMRGIRGADFOCQQARAVLAGTFRAGLFLRSRQLDLYSTI 60  
 Db 1 hshrdfqpvihivalnsplsggmrgrgadfqcfqgaravlagtfragraffslridlysi 60

OY 61 VRRADRAAVPTVNLKDELLEPSWEALFSGSGCPKPGARTFSFDGKVLRHTPQKSWV 120  
 Db 61 viradravlpvnikdellepswealfsgsgcpkpgarifsfqdgkvrlhtpwpksvw 120

OY 121 HGSDPNGRRLTESYCERWRTEAPSATQASSLLGRGLGOSAASCHAYIVLCIENSFMT 180  
 Db 121 hgsdpngrrltesycetwrteapsatqassllgrglgosaaschayivlciensfmt 180

OY 181 AS 182  
 Db 181 as 182